Faculdade de Engenharia da Universidade do Porto
Departamento de Engenharia Mecânica

Analysis of failures of rolling stock railways rolling bearings

Hermínio Maio Graça Fernandes

Dissertation in Mestrado Integrado em Engenharia Mecânica

Advisor:
Prof. Luís António de Andrade Ferreira

Porto, February 2017

Abstract

Railways companies are under great pressure to deliver more while providing an affordable, safe and high quality service. Asset management is a valuable tool for these companies to keep up with the demand. It focus on different areas within which resides the monitoring of assets condition. One of most important assets in railway companies is the rolling stock. Railway vehicles are composed of various components, that inevitably fail. The axle-bearing, responsible for the overall performance and safety, is one of the most important components. However, it has the higher failure rate, and so its monitoring is of great importance. One of the best monitoring techniques in regard to cost effectiveness and safety is condition monitoring. It is quite a vast topic, and for each type of monitoring, the selected procedure should be chosen wisely. Despite the attention given recently to condition monitoring, the monitoring of slow-speed bearings is quite an unexplored area.

The present work reviews briefly the overall information about the existing bearings and railways components. The importance of asset management in railway organizations is also discussed. Condition monitoring basis are described, as well as several techniques for data collection, process and analysis. Some results regarding the use of an existing test bench, on the Faculdade de Engenharia da Universidade do Porto (FEUP), are also displayed in this document. It was intended to study the behaviour of the SKF TBU 1639590 A, used in the Comboios de Portugal (CP), UTE 2240 unit. Its functioning was simulated in the test bench, for different loads and speeds, and the results were recorded and analysed.

Resumo

Nos dias de hoje, é exigido às empresas ferroviárias um serviço sucessivamente mais rápido e seguro a preços cada vez mais baixos. A gestão de activos tem um papel fundamental na resposta positiva a estas exigências, e actua em diversas áreas, entre as quais se destaca a monitorização da condição dos activos. Um dos activos mais importante das empresas ferroviárias é o seu material circulante. Os veículos ferroviários são compostos por vários componentes que irão inevitavelmente falhar. O rolamento das caixas de eixo é um dos componentes mais importantes, assegurando a *performance* e segurança do veículo. No entanto, é também o componente que mais vezes falha, sendo fundamental, portanto, a sua monitorização. Uma das técnicas de manutenção mais seguras e economicamente eficientes é a Condição de Monitorização. Existem diversas técnicas para a monitorização de condição, sendo que, geralmente, existe uma mais indicada para cada tipo de activo a ser monitorizado. Apesar da importância que lhe tem sido dada ultimamente, a monitorização de rolamentos de baixa velocidade é um tópico que tem recebido pouca atenção.

Este trabalho apresenta uma revisão acerca de conceitos básicos sobre o funcionamento dos rolamentos e componentes dos veículos ferroviários. É também abordada a importância da gestão de activos nas empresas ferroviárias. São descritos os aspectos fundamentais da condição de monitorização, assim como técnicas para a recolha, processamento e análise dos dados obtidos por esta via. Apresentam-se ainda os resultados obtidos no banco de ensaios existente na Faculdade de Engenharia da Universidade do Porto (FEUP), onde foi estudado o comportamento do rolamento cónico da SKF, TBU 1639590 A, usado pela empresa Comboios de Portugal (CP), nos veículos das UTE 2240. Com o banco de ensaios existente, simulou-se o funcionamento do rolamento a várias velocidades e a diferentes cargas aplicadas, tendo sido recolhido e analisado os diferentes resultados.

Agradecimentos

Nenhum indivíduo é detentor da verdade ou do conhecimento absoluto, pelo que qualquer tarefa realizada será sempre influenciada pelas experiências e contactos a que o sujeito se vai submetendo.

Gostaria de agradecer, em primeiro lugar, ao Prof. Luís Andrade Ferreira, pela orientação e acompanhamento do trabalho realizado, por todas as sugestões e correcções, e por me ter dado as bases para que este trabalho fosse possível. Um agradecimento especial ao Prof. José Dias Rodrigues pelo acompanhamento constante, e pela resposta positiva aos pedidos de ajuda sucessivos, mesmo que inoportunos. Ao Engº Ricardo Carvalho por, de um momento para o outro, se ter disponibilizado a ajudar-me na montagem do amplificador da célula de carga. Ao Diogo Amorim e à Safa Arab, por me terem recebido no seu local de trabalho. Ainda à Faculdade de Engenharia da Universidade do Porto, pela oportunidade desta etapa que agora termina.

Um agradecimento especial à minha família, que me moldou e me acompanhou desde o início. Aos meus Pais, por todo o apoio incondicional e pelos valores transmitidos, pela pessoa que formaram e pelas bases que lhe deram para ser feliz e para se realizar. Aos meus irmãos, Manuel e Manuela, pela forma como receberam o seu irmão mais novo, pela nossa relação e pelas pessoas que são. À Babá pelas tardes, passeios, conversas e momentos que passamos juntos e que espero que nunca acabem. À Tia Milai, a madrinha de todas as horas. Foi único crescer convosco, não seria possível escolher outra família!

À Leonor, ao Rui, ao Zeferino e ao Roda, pela sua amizade e por me terem acompanhado nos primeiros anos do curso, por me terem mostrado que a faculdade valia a pena, e por me mostrarem o Porto.

À Diana, à Filipa, ao Tiago, ao André, à Sofia e ao Diogo, pela amizade incondicional mesmo nos tempos mais difíceis e pelo grupo de amigos que foram e serão. Pelas manhãs, tardes e noites mal dormidas, pelos vários projectos em que embarcamos, e por mostrarmos que vale a pena lutar por aquilo em que acreditamos.
À Clarisse, um beijinho especial pelo carinho demonstrado.

Por fim, ao Orfeão Universitário do Porto por tudo aquilo que me ensinou e pela formação que me deu *fora da faculdade*. Por me ensinar a valorizar a arte e a cultura Portuguesa, por me ensinar a dançar, cantar e estar em palco. Por me ensinar a dirigir, organizar, realizar (e tantas outras aptidões) e me dar as bases para ser um bom profissional. Porque todo o tempo "investido" valeu a pena, e porque me deu muito mais do que algum dia conseguirei retribuir.
A todos os que já por lá passaram, um obrigado!

Contents

1 Introduction **1**
- 1.1 Introduction . . . 1
- 1.2 Objectives . . . 2
- 1.3 Structure . . . 3

2 Rolling bearings **5**
- 2.1 Types of rolling bearings . . . 5
  - 2.1.1 Ball bearings . . . 6
  - 2.1.2 Roller bearings . . . 7
  - 2.1.3 Thrust roller bearings . . . 9
- 2.2 Rolling bearings failure . . . 10
  - 2.2.1 Fatigue . . . 11
  - 2.2.2 Wear . . . 11
  - 2.2.3 Corrosion . . . 13
  - 2.2.4 Electrical Corrosion . . . 14
  - 2.2.5 Plastic Deformation . . . 15
  - 2.2.6 Fracture and cracking . . . 16
- 2.3 Bearings characteristic frequencies . . . 17

3 Railway **21**
- 3.1 Bogies . . . 21
  - 3.1.1 Types of bogies . . . 22
  - 3.1.2 Bogies components . . . 24
- 3.2 Axle-bearing for railway applications . . . 25
- 3.3 Types of axle failures . . . 26
- 3.4 Asset management in railway organizations . . . 34
  - 3.4.1 Asset polices . . . 36
  - 3.4.2 Condition monitoring in railway organisations . . . 37
- 3.5 Conclusions . . . 38

4 Condition Monitoring **41**
- 4.1 Maintenance techniques . . . 41
  - 4.1.1 Condition based maintenance versus Time based maintenance . . . 42
- 4.2 Condition based maintenance . . . 44
  - 4.2.1 Condition based maintenance procedure . . . 45
- 4.3 Data Process . . . 47
  - 4.3.1 Signal Processing Techniques . . . 47
  - 4.3.2 Denoising techniques examples . . . 54
- 4.4 Data Analysis . . . 54
  - 4.4.1 Failure Diagnostic Techniques . . . 54
  - 4.4.2 Failure Prognostic Techniques . . . 56

5 Experimental test **61**
5.1 Test bench . . . 63
5.2 Acquisition of data . . . 64
5.2.1 Force acquisition equipment . . . 64
5.2.2 Vibration acquisition equipment . . . 66
5.2.3 Data acquisition . . . 67
5.3 Signal Analysis . . . 67
5.3.1 Statistical Parameters . . . 69
5.3.2 Fast Fourier transform . . . 71
5.3.3 Spectral Kurtosis and Kurtogram . . . 74
5.4 Validation of results . . . 76
5.4.1 Acquisition system . . . 76
5.4.2 Fast Fourier Transform . . . 78
5.4.3 Spectral Kurtosis and Kurtogram . . . 79
5.5 Previous calculations . . . 80
5.5.1 Calculation of dynamic bearing loads . . . 81
5.5.2 Calculation of bearing characteristic frequencies . . . 83
5.6 Test bench results . . . 85
5.6.1 Measurement complications . . . 85
5.6.2 Obtained measurements . . . 85
5.7 Future alterations to the existing equipment . . . 88

6 Conclusion **89**
6.1 Conclusions . . . 89
6.2 Future developments . . . 90

A Axle bearing **95**

B UTE 2240 **97**

C MATLAB code **99**

List of Figures

2.1 Rolling element bearing components (a ball bearing is depicted). 6
2.2 A deep-groove ball bearing (one row) on the left, and a double-row deep-groove ball bearing on the right [3]. 6
2.3 A angular-contact ball bearing (one row) on the left, and a double-row angular-contact ball bearing on the right [3]. 7
2.4 A cylindrical roller bearing (one row) [3]. 8
2.5 A needle roller bearing (one row) [3]. 8
2.6 A tapered roller bearing (one row) with a detail of its contact angle, on the left, and a two-row tapered roller bearing on the right [3]. 8
2.7 A spherical roller bearing (one row) [3]. 9
2.8 From left to right: spherical, cylindrical, tapered and needle rollers thrust bearings [3]. 10
2.9 Example of sub-surface initiated fatigue (on the left) and surface initiated fatigue (on the right) on a bearing outer ring raceway [6]. 12
2.10 Example of abrasive wear on a steel cage of a tapered bearing (on the left) and adhesive wear on the outer inner ring raceway of a cylindrical roller bearing (on the right) [6]. 13
2.11 Example of moisture corrosion on the inner ring of a spherical roller bearing (on the left), fretting corrosion in the bore of a bearing (on the center) and false brinelling damage in the outer ring of a tapered roller bearing (on the right) [6]. 13
2.12 Example of excessive voltage on a ball bearing raceway (on the left) and wash-boarding in a raceway of a deep-groove ball bearing (on the right) [6]. 14
2.13 Example of overload from wrong mounting (on the left), identation from debris of soft and hard particles (on the center) and indentation from handling on a inner ring of a cylindrical roller (on the right) [6]. 15
2.14 Example of forced fracture on a bearings outer ring provoked by a bad housing seat (on the left), fatigue fracture of a spherical roller bearing outer ring (on the center) and thermal cracking resulting on transverse cracks on the side face of a bearing inner ring (on the right) [6]. 17
2.15 Typical vibration signals and envelope signals from local faults in rolling ellement bearings [3]. 18

3.1 Circumscription of the carriage with the track, achieved through the use of a bogie [11]. 22
3.2 Single-axle and two-axle bogie and the comparison of of effect of track irregularity [12]. 22
3.3 Articulated and non-articulated bogie [12]. 23
3.4 Bolster and bolsterless bogie [12]. 23
3.5 Swing hanger and small lateral stiffness bolster spring bogie [12]. 24
3.6 Components of a bogie (DT50 Bogie model) [12]. 24

3.7 Damage on a bearing that heated to unacceptable temperatures due to loose bearing components [13]. 27
3.8 Sign of water ingress (on the left and center) and raceway etching (on the right) are signs of water etch failure [13]. 27
3.9 Whell tread defects (on the left) usually leads to bearaing failure as, for example, breaking of bearings cage (on the center and right) [13]. 28
3.10 Examples of bearing distress caused by fatigue spalling [13]. 28
3.11 Examples of bearing distress with unkown causes [13]. 29
3.12 Signs of external abuse that lead to bearing distress [13]. 29
3.13 Bearing distress due to lubrification break down [13]. 30
3.14 Bearing distress (on the left) caused by displacement of the adapter, which is represented on the figure on the right [13]. 31
3.15 Displaced seals and consequential bearing distress [13]. 31
3.16 Bearing distress caused by uneven loading from truck components [13]. 32
3.17 An incorrect assembly procedure (on the left) can cause bearing distress (on the right)[13]. 32
3.18 An improper manufactiring (bearing dimensions being checked on the left image) can cause bearing distress (signs of etch on the rollers, on the left and middle images) [13]. 32
3.19 Asset Management Consulting Ltd's asset management model[15]. 36

4.1 Cost rate under optimal TBM strategy and CBM strategy. 43
4.2 Classification system of fialure prognostic techniques [37]. 56

5.1 A photo of the UTE 2240 (yellow model). 61
5.2 Model of tapered rolling bearing unit (TBU) similar to the rolling bearing TBU 1639590 A (130 × 220 × 150) [6]. 62
5.3 Photograph showing the existing test bench. 63
5.4 Schematics of the test bench. 63
5.5 Access to the outer race of the bearing. 64
5.6 Kistler ICAM 5073A. 64
5.7 Cables connected to the D-Sub 9-pin and D-Sub 15-pin connectors. 65
5.8 *ManuWare* software window's aspect during a regular measurement. 66
5.9 Flowchart indicating the main steps of the signal analysis process adopted. . . . 69
5.10 Matrix representation of the Discrete Fourier transform [42]. 72
5.11 Matrix representation of the Fast Fourier transform [42]. 72
5.12 Diverse Windowing techniques that can be used for different types of signals [43]. 73
5.13 Montage used for the validation of the acquisition's system results. 76
5.14 Results of different acquisitions. 77
5.15 Time domain representation of the signal. 78
5.16 Frequency domain representation of the signal. 78
5.17 Windowing technique (Hanning) applied to the signal. 79
5.18 Resulting Kurtogram 79
5.19 Comparison of the original measured signal to various envelopes. 80
5.20 Comparison of the signal's power spectral density and the result obtained using *psde.m*. 80
5.21 Power spectrum of several envelopes, which bandwidths were chosen according to the kurtogram. 80
5.22 Considered acting bearing loads [6]. 81
5.23 Exemple of symmetrical axleboxes designs (on the left), of non-symmetrical axleboxes designs (on the center) and a picture of the axlebox of the UTE 2240, symmetrical(on the right) 83

5.24 Schematic representation of value l, the distance between 2 load center of a double-row tapered roller bearing [6]. 83
5.25 Experimental test setup. 86
5.26 Measured values (time-domain) at different speeds. 86
5.27 Measured values (time-domain), at 397 rpm, for two different loads. 87
5.28 Measured values (frequency) at different speeds. 87
5.29 Measured values (frequency-domain), at 397 rpm, for two different loads. 87

A.1 Technical drawing of the *TBU 1639590 A* bearing. 95

B.1 Technical drawing of UTE 2240 (part 1) 97
B.2 Technical drawing of UTE 2240 (parte 2) 98
B.3 Technical drawing of UTE 2240 98

List of Tables

3.1 Axle failure and the corresponding bearing failures. 33
3.1 (continued) 34

4.1 The four key steps of the condition based maintenance procedure. 47
4.2 Resume of the signal processing techniques covered in this chapter. 48
4.3 Type of failure diagnostic techniques 55
4.4 Failure prognostic techniques covered in this chapter 57

5.1 Pins allocation and corresponding cable colour for the D-Sub 15-pin interface. . . 65
5.2 Parameter setting that is common to both scripts 67
5.3 Continuous acquisition details 68
5.4 Non-continuous acquisition details 69
5.5 Principal lines of code for the calculation and representation of the frequency domain. 75
5.6 UTE's considered weight, in Kg 81
5.7 Values used in the calculation of maximum static axlebox load, G, in KN 81
5.8 Values used in the calculation of mean radial load, K_r (G and K_r in KN) 82
5.9 Values used in the calculation of mean radial load, K_a (G and K_a in KN) 82
5.10 Values used in the calculation of mean radial load, F_r (D_a and l in mm and K_r, K_a and F_r in KN) 83
5.11 Values of radial and axial bearing loads, K_r and K_a, respectively, in KN 83
5.12 Range of shaft frequencies considered 84
5.13 Values of bearing characteristic frequencies, in [Hz], for various values of train velocity, in [km/h] 84
5.14 Values of bearing characteristic frequencies, in [Hz], for various values of train velocity, in [km/h], obtained in the SKF's frequency calculator tool 85

Chapter 1

Introduction

1.1 Introduction

Since the dawn of organized society, the proper transportation of people and goods has been of huge importance to its development and economic growth. The search for a quick and safe transportation, at low cost, has been a constant goal.

Several transportation methods were developed and used through the centuries (navigation, railway, automobile, airplane), and have evolved to their present form. Railway, as we know it today, appeared for the first time in the 19th century, in British mines. They had the form of simple carts that were guided by the movement of the wheels on the tracks. Since this day, huge developments were made, both to the composition of the overall structure and its locomotion method. With the industrial revolution, steam power was used, contributing to the widespread of tracks and the achievement of considerable high speed vehicles, that were further increased in the beginning of the 20th century with the adoption of the electric traction.

In the second half of the past century, the last developments have taken things a big step forwards. The use of automatic train control, Geographic Positioning System (GPS), Intelligence Techniques, technical innovations and the automation of most components have permitted the railway to improve their competitive position in the transport market.

Today, the railway companies are under great pressure to keep up with the transportation alternatives as airplains, passenger cars, buses, trucks and sea transport. There were big demands for the modernization and improvement of the railway industry, that could no longer be kept up with just mechanical innovations. Although the pursuit of improvements for higher speeds never ceases, managerial actions have assumed a huge importance. Solution in organization, management, costs and supply of service must also be seek.

Asset management is receiving great attention from the railway companies. It is defined by the British Standards Institute Publicly Available Specification for the management of physical assets (BSI PAS 55-1) as the *Systematic and coordinated activities and practice through which an organization optimally manages its assets and their associated performance, risks and expenditures over their lifecycle for the purpose of delivering the organization's business objectives.* It can focus on quite diverse areas as: establishment of strategic plans, study of the assets requirements and performance, prioritisation of investment, use of management techniques, and monitoring of the assets performance and condition.

The establishment of a proper maintenance wouldn't be possible if not for asset polices. Among other things, asset polices defines the maintenance planning based on the minimization of whole life and system cost focusing on reliability, availability, maintainability and safety specification.

On of the most cost effective techniques nowadays is condition monitoring. It provides information about a component's or machine's state, obtained from embedded sensor or external tests and measurements. With the late development of technology, the access to measuring

equipment at affordable prices and the use of equipment with higher computation power, condition monitoring is being adopted by many companies to monitor expensive machinery and critical components.

Roller element bearings are one of the most used components in machinery and are crucial for the functioning of every rotating machine. Their overall function is hardly replaced by any other component, as they provide several advantages with a considerably simpler mechanical composition and lower cost than other more complex alternatives. However, they are a component with one of the highest failure rate that hardly reach their service life.

Comboios de Portugal (CP) is one of the oldest and biggest railway companies of Portugal which operate freight and passenger trains in Portugal. It owns several rolling stock units, and is responsible for their management. They have showed some concern towards the maintenance plan applied to one of its assets, the UTE 2240 (a iberian gauge electric multiple units with three cars) and, in particular, towards the maintenance plan applied to its axle bearings. It lies only on the replacement of the bearing when a certain service period is reached.
Each cart has a total of 8 bearings, having the unit a total of 24. The bearings used, constructed specially for railway application, are designed to withstand harsh condition and were build according to strict standards, resulting in a high cost. The replacement of every bearing represent a huge expense, specially if it is borne in mind that they are being replaced regardless of their health state.

The main objective of the present work is to study the possibility of using condition monitoring on these bearings and develop techniques and experimental procedures for its implementation. For this purpose, it was used some already existing equipment in the Faculdade de Engenharia da Universidade do Porto (FEUP). A test bench was designed and constructed in the year of 2011 and 2012, and *National Instrument*'s equipment was bought also in 2012, integrated in the SIEF Project (Sistema Integrado da Fiabilidade de Equipamentos Ferroviários). The existing information about these equipments, their montage and usage for data acquisition is scarce, so further developments must be done in this area. A process that allows the acquisition of the measured values, their treatment and posterior analysis must be developed.

It is expected to study the normal functioning of the bearing and the associated behaviour, through the collection of several vibration features. Also, it is intended to simulate various bearing defects in the bearing, and compare the results to the healthy bearing's ones.

1.2 Objectives

- Review the existing bearing types, their deterioration evolution and related failures.
- Review the existing type of mechanical structures used for railway applications, and the location of their main components.
- Study the most frequent failures in a railway vehicle axle-system and link it to the reviewed bearing failure.
- Understand the importance of condition monitoring in railway companies.
- Understand the fundamentals of condition monitoring. Study its different techniques and the state of the technology.
- Decide what techniques should be applied for the vibration monitoring of the axle-bearing from the CP's UTE 2240.
- Design and test an acquisition system to be applied to the existing test bench, using the available acquisition equipment. Build and program an application to enable the collection of the measured signal by a computer. Also, it should allow the treatment of the signal and output of results of interest.

- Collect different features extracted from the acquired result, and study their evolution with the rotational speed and the load applied.

1.3 Structure

This document is divided in six chapters.
The following chapter addresses itself to review the importance of rolling bearings and the existing bearing types. Various bearing faults are described, and it is studied how the bearing responds to them.
The next chapter (chapter 3) is subdivided in several sections. The first sections are related to the mechanical components used in railway vehicles, their location and main features. The most common axle-failures are reviewed and linked to the previously studied bearing failures. The following two section review the importance of condition monitoring in railway applications and conclude with a consideration linking the last two chapters.
Chapter 4 clear the fundamental aspects on maintenance techniques and condition monitoring. The main steps of condition monitoring are described, and several different techniques for the acquisition, treatment and analysis of the signal are presented.
Chapter 5 describes the experimental setup used, describing the existing material and how it should be connected. Then, detailed information about the techniques used for the acquisition and treatment of the signal is mentioned. The *MATLAB* code created is described in this section to ease further alteration or adaptation in the future. Data needed for the experimental setup is calculated, and the obtained result are displayed and commented.
The last chapter, Chapter 6, contains the conclusions of this work, and future developments are recommended.

Chapter 2

Rolling bearings

Bearings are widely used in all kinds of machinery as they provide the transmission of forces between the moving part and the main structure at a low friction[1]. Only rotating bearings will be described, as linear bearings (designed for linear movement) lie outside the scope of this report.

There are various types of bearings with different denominations, such as gas film bearings, foil bearings, magnetic bearings, externally pressurized hydrostatic bearings, rolling element bearings, etc, each with its specialized field of application.
Hydrostatic bearings are utilized in applications where ample supply of pressurized fluid is available and extreme rigidity under heavy loading is required. Self-acting gas bearings are used in applications where the loads are light, speeds are high, a gaseous atmosphere exists, and friction must be minimal. Rolling element bearings proves to be the type that meets the requirements of almost every kind of application, for example, cases where the loads and temperature are high, the environment is dusty and dirty, there are extreme changes in temperature and vacuum environments. They are usually the first choice in life-critical applications. Their utilization has several advantages as[2]:

- They operate with less torque friction than hydrodynamic bearings, resulting in smaller power losses and friction heat generation;
- They are less sensitive to load fluctuations than the hydrodynamic bearings;
- They require small amounts of lubricant and have the potential possibility of operating with a life-long supply of lubricant;
- If properly design, they can support simultaneous radial and thrust loads;
- They provide a wide load-speed range;
- They are relatively insensitive to fluctuation in load, speed and operating temperature.

The main disadvantage of this type of bearings is that they eventually fail due to fatigue of the surface in rolling elements, independently of its lubrication, mounting and conditioning procedures.

2.1 Types of rolling bearings

Rolling bearings have four main components: an inner and outer race, the rolling elements and the cage. The rolling elements are hold between the two races, and their non-contact and evenly distribution is assured by the cage. Additional components can be present, as seals or shields on both ends, protecting the bearing from the entrance of external particles, and spacers, that reduce the wear on the side faces of the races.

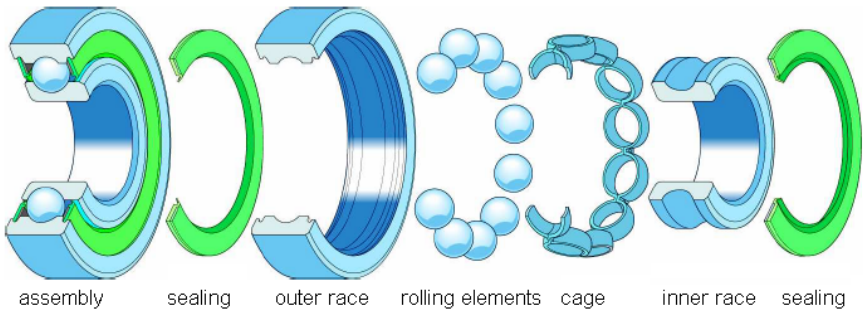


Figure 2.1: Rolling element bearing components (a ball bearing is depicted).

According to the format of the rolling elements, the bearing can be divided in two main types, ball bearings and roller bearings, and various sub-types. A short description of each sub-type can be found in [2].

2.1.1 Ball bearings

Ball bearings rolling elements have a spheric shape. They are widely used as they can support high-loads and have the lowest cost when compared to other kind of bearings.

Radial ball bearings - Deep-groove ball bearing

Single-row deep-groove ball bearings are the most popular type. Their main features makes them a good and cost-effective choice on most applications. They have high load-carrying capacity if equipped with contaminant-exclusion equipment and correct lubrication.

Although they were designed to carry radial load, they can support combined radial and thrust loads, and even withstand misaligning of small magnitude if they feature a proper cage design. Greater radial-loads can be supported if a double-row deep-groove ball bearing is used. With adequate lubrication and cooling, they can be used in high speed applications.

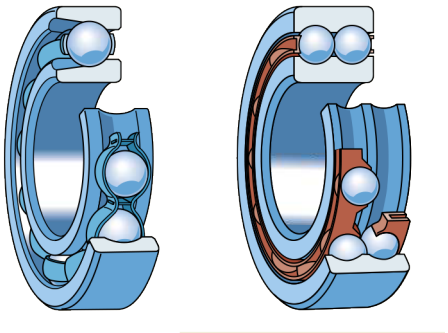


Figure 2.2: A deep-groove ball bearing (one row) on the left, and a double-row deep-groove ball bearing on the right [3].

Radial ball bearings - Angular-contact ball bearing

Unlike deep-groove ball bearing, in which the contact is made around the perpendicular direction of the axle, angular-contact ball bearings have a contact angle that usually doesn't exceed 40^{o}.

They can support combined radial and thrust load. The contact angle is proportional to the magnitude of the supported thrust load. Arranging them in a tandem disposition results

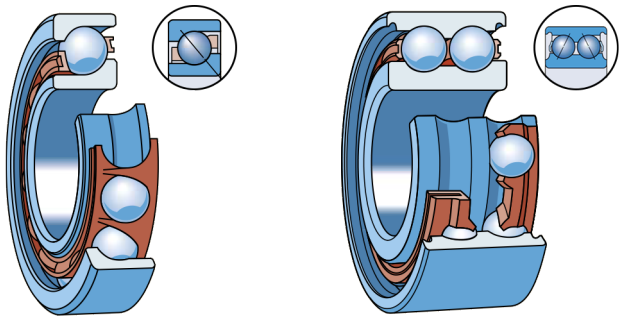


Figure 2.3: A angular-contact ball bearing (one row) on the left, and a double-row angular-contact ball bearing on the right [3].

in a greater thrust-carrying capacity. Preloading the pair will stiffen the assembly in the axial direction.

Double-row angular-contact bearings can carry a combination of thrust and radial loads, or a thrust load in both directions. Using larger roller balls it is possible to produce self aligning bearings.

2.1.2 Roller bearings

Roller bearings are used in application that can't be solved by the utilization of ball bearing assemblies. Although they cost more to manufacture when compared to ball bearing assemblies, they offer a much stiffer structure and have a large (radial and axial) load-supporting capability. They require great care in mounting.

Self-aligned spherical roller bearings are an excellent choice to solve problems of accuracy of alignment of shafts and housings.

The main difference between this kind of bearings and the ball bearings is the shape of the rolling elements, which can be one of the following: cylindrical, conical or a cylinder with a curved generatrix.

Radial roller bearings - Cylindrical roller bearing

This type of bearing has an exceptionally low-friction torque characteristic and high radial-load carrying capacity, thus proving ideal for high-speed operations. For even greater radial-load-carrying capacities, it is preferable to use bearings composed of two or more rows of rollers than using bearings with longer rollers, as the tendency of the rollers to skew is smaller in the first case.

The typical cylindrical bearing is free to float axially but, if equipped with a guide flange, it can support thrust loads.

Radial roller bearings - Needle roller bearing

The needle rollers have considerably larger length than diameter. The use of this kind of bearing is ideal for applications in which the radial space is on a premium, when an oscillatory motion occurs or when a continuous rotation occurs at a light and intermittent load. In some applications it can be even mounted directly on a hardened shaft.

Needle roller bearings produce great friction when compared to other cylindrical roller bearings, because their manufacture is not very accurate and so they cannot be guided as well as other bearings.

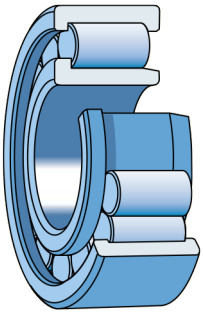


Figure 2.4: A cylindrical roller bearing (one row) [3].

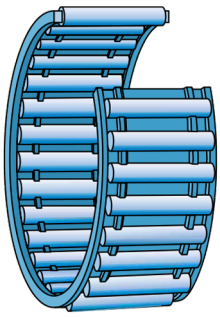


Figure 2.5: A needle roller bearing (one row) [3].

Tapered roller bearings

Tapered rollers have a steep angle that varies from bearing to bearing. The inner ring is called the cone and the outer ring the cup.

This type of bearing is specially useful in applications that combine large radial and thrust loads, although they can carry thrust loads only. One-row bearings are mounted in pairs: this provides several mechanical benefits, as it simplifies the mounting procedure, enabling one bearing to be adjusted against the other. Two-row bearings prove themselves useful in applications were greater radial loads are present, eliminating problems of axial adjustment. Other combinations can be used: pairs of double-row or four-row bearings.

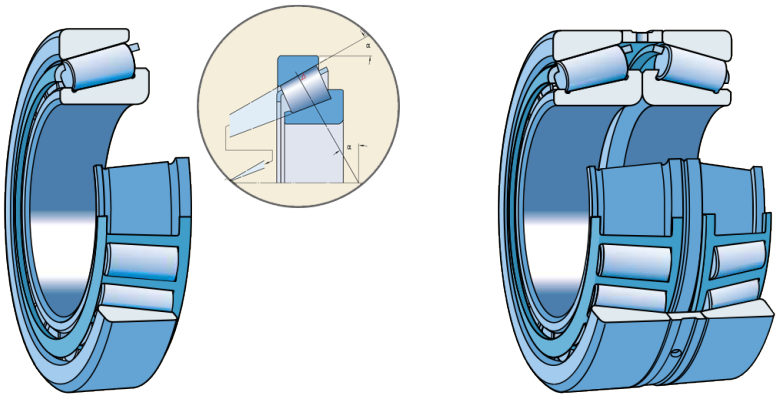


Figure 2.6: A tapered roller bearing (one row) with a detail of its contact angle, on the left, and a two-row tapered roller bearing on the right [3].

Spherical roller bearings

Spherical roller bearings have an outer raceway shaped like a portion of a sphere and rollers with a curved generatrix in the transverse direction of rotation that conform closely to the inner and outer race. This design make the bearing internally self-aligning and able to support high loads.

Double-row bearings can be used to carry combined radial and thrust loads. The load-carrying capacity of a conventional spherical bearing can be increased by using a double-row bearing with barrel shape and symmetrical rollers, eliminating undercuts in the inner raceway allowing the use of longer rollers.

These roller bearings have greater friction than cylindrical roller bearings, due to the sliding of the roller, proving not to be suitable for high-speed applications. As expected, they can't support moment loading.

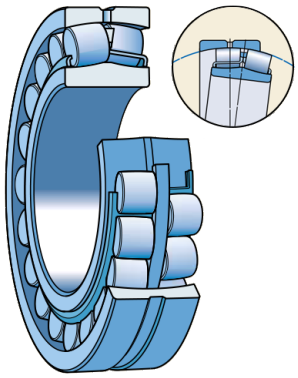


Figure 2.7: A spherical roller bearing (one row) [3].

2.1.3 Thrust roller bearings

Thrust roller bearings are mainly used in applications with vertical axis. All the subtypes of thrust bearings can carry axial loads, but each with its own specifications.

Spherical rollers thrust bearings have high-load-carrying capacity and can carry a combination of thrust and radial load. They are not suitable for high-speed operations. They can accommodate heavy axial and radial loads.

Cylindrical rollers thrust bearings experience a large amount of sliding that can be reduced with the use of multiple short rollers. They are limited to low-speed operations. They can accommodate heavy axial loads and shock loads, but must not be subjected to radial loads.

Tapered rollers thrust bearings generate a considerable sliding frictional force. They are almost limited to relatively low-speed operations. They can carry heavy axial loads, they are stiff and insensitive to shock loads.

Needle rollers thrust bearings are similar to cylindrical roller bearings and are useful in applications with a narrow axial space. As the sliding is prevalent to a greater degree, the loading must be light. The can carry heavy axial loads and shock loads.

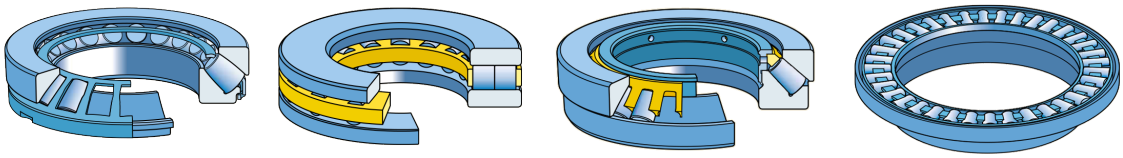


Figure 2.8: From left to right: spherical, cylindrical, tapered and needle rollers thrust bearings [3].

2.2 Rolling bearings failure

A bearing is specially chosen to operate under certain conditions, but it eventually fails if used for long enough. Several mechanisms affect the bearing integrity.

Rolling contact fatigue has the most prominent effect in the degradation of the bearing, propagating from surface and subsurface initiated failures: when the applied stresses exceed the elastic limit of the bearing steel, microstructural changes occur. Cracks are initiated at subsurface flaws of the race, eventually leading to macro pitting due to the passage of rolling element. Small "spalls" of material are released inside the bearing affecting the correct functioning of the bearing. The rolling bearing will eventually reach such a disrupted state that its utilization is no longer acceptable and its replacement is crucial.[4]

Aside from the wear mechanisms, there can be several stress concentration mechanisms (most of them caused by the wear mechanisms themselves) that can affect the bearing behaviour. Subsurface inclusions, asperities, dents and debris (as moving asperities) are some of the given examples and are the most significant mechanisms. Subsurface inclusions, in conjunction with cyclic stress, tend to provoke the initiation of cracks that latter propagate until they reach the surface. Asperities can be expressed as a form of surface roughness, being the most significant originated by dents, abraded points or ploughed points. A dent can be described as a two-dimensional stress raiser (of two up-ward asperities), contributing to the initiation of cracks. The small material, detach by growing cracks, act as minor indenters once they are hard and sharp enough. Their action can be reduced by adhesive and abrasive processes. The zone from were material has detach will tend to increase and initiate new asperities [5].

The period of time until the moment when the bearing has reached its limit and shall no longer operate is known as life expectancy. It is calculated based on various assumptions as: the type of bearing; the quality of bearing material an potential defects; the dimensions of the bearing and their relation to the other components; the mounting procedure; the quantity and type of lubricant; the type of existent protection components, as the existence of seal; the operating conditions; the existence of maintenance.[6]

Unfortunately, the life expectancy calculated is quite hypothetical, as the real operating conditions often diverge from the ideal ones. According to [7], rotating machines have the highest failure rate and only 20% of rolling bearing reach its service life. The smallest unpredictable alteration may provoke damages on the bearing that will shorten its life expectancy. So, it is importante to know what type of damage exists, and how it affect the service life of the bearing. International Organization for Standardization (ISO) has developed a classification system that divide the damage in six different modes and sub-modes, related to post-manufacturing sustained damage, that will be described below according to [6]. The visual inspection of the bearing and its surfaces allow the identification of almost every kind of damage and the mechanism involved.

2.2.1 Fatigue

The most usual symptoms of fatigue is the breaking out of material, as spalling or flaking. It is caused by repeated stresses developed in the contact of the different components of the bearing that changes the structure of the material.

Sub-surface initiated fatigue

Aspect	It is characterized by cracks and micro-cracks that are develop underneath the raceway surface. As they grow and reach the surface, the material tend to break loose and spalls occur.
Causes	It usually occurs after very long running time, but it can occurs earlier by a heavy load due to abnormal operating conditions.
Acceptability	A bearing showing damage of this kind is no longer acceptable and should be considered scrap.
Solutions	Check bearing seats for conformity and adjust loading conditions.
Appearance	It often occurs in the shaft and housing seats, but usually only after a very long running time.

Surface initiated fatigue

Aspect	It is characterized by the plastic deformation of contacting surfaces. Zones of spall and wear are observable.
Causes	It is usually caused by inadequate lubrication, that don't permit a proper lubricant film. The asperities of each surface touch, and shear over each other. The asperities break, micro-spalls occurs and starts growing to larger sizes, leading to a combination of spalling and wear around the load zone. It can occur due to the supply of a wrong kind of lubricant or due to the contamination of the existing one.
Acceptability	A bearing showing damage of this kind is no longer acceptable and should be considered scrap.
Solutions	Choose an adequate lubricant (check the appropriate type and quantity of grease), change the replenishment interval, replace the existing seals.
Appearance	It usually occurs in load zones and inner ring.

2.2.2 Wear

It is characterized by the loss of material due to interaction of the asperities of contacting surfaces during service.

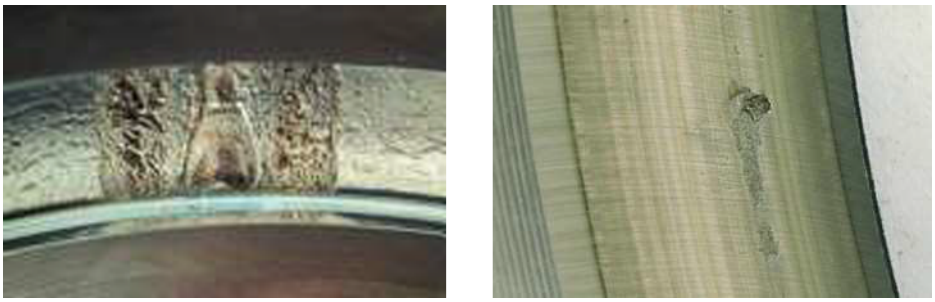


Figure 2.9: Example of sub-surface initiated fatigue (on the left) and surface initiated fatigue (on the right) on a bearing outer ring raceway [6].

Abrasive wear

Aspect	Depending on the abrasive particles involved on the process, it can have two opposite appearances. Most of the times it is characterized by the appearance of dull surfaces, but when the particles involved act like polishing material, the surfaces become extremely shiny.
Causes	It usually caused by inadequate lubricant, usually because of the ingress of abrasive contaminant particles.
Acceptability	When the surface presents a mirror finish, the bearing components are still acceptable if certain clean condition are met. However, if it presents ridges that can be felt by fingernails or other blunt probes, the components should be scrapped. It is an accelerating process: as the number of wear particles increase, the lubricant properties worsen and vice versa. Corrective measure should be taken early when the problem is detected.
Solutions	Use polymer cages, replace seals or lubricant, rethink the lubricant's grease type.
Appearance	The cage is usually the critical component, were the signs of wear will predominate.

Adhesive wear/Smearing

Appearance	It is characterized by the transference of material from one of the contacting surfaces to the other, due to friction heat. Sometimes, it causes the tempering or rehardening of the surface.
Causes	It is usually caused by inadequate lubrication, that don't permit a proper lubricant film. It can occur due to the supply of a wrong kind of lubricant or due to the contamination of the existing one.
Acceptability	A bearing showing damage of this kind that is detectable by drawing the fingernail across the damage is no longer acceptable and should be considered scrap.
Solutions	Choose an adequate lubricant (check the appropriate type and quantity of grease), change the replenishment interval, replace the existing seals.

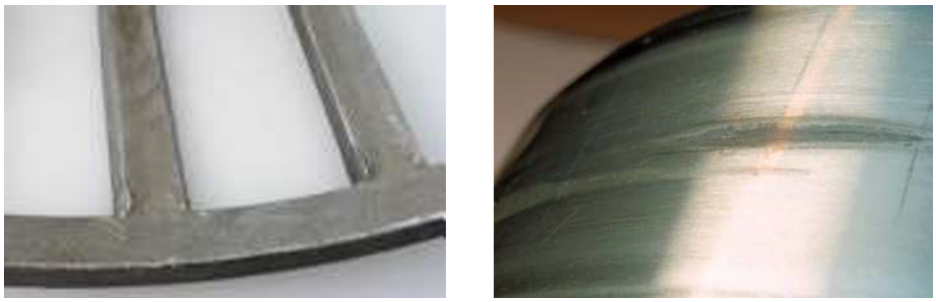


Figure 2.10: Example of abrasive wear on a steel cage of a tapered bearing (on the left) and adhesive wear on the outer inner ring raceway of a cylindrical roller bearing (on the right) [6].

2.2.3 Corrosion

It is characterized by the oxidation of metal in the bearing by an abnormal contact of the inside surfaces with water or air.

Figure 2.11: Example of moisture corrosion on the inner ring of a spherical roller bearing (on the left), fretting corrosion in the bore of a bearing (on the center) and false brinelling damage in the outer ring of a tapered roller bearing (on the right) [6].

Moisture corrosion

Aspect	Components suffering from this kind of damage will present greyish black streaks corresponding to the rolling element distance and may contain formations of deep-seated rust.
Causes	It is caused by the presence of water or corrosive agents in the inside of bearing in such quantities that the lubricant can no longer provide an adequate protection to the steel surfaces.
Acceptability	If the damage is removed by polishing with a fine abrasive paper, it can be used, otherwise, or if it is felt with a finger nail, it should be considered scrap.
Solutions	Replacement of seals, rethink the lubricant grease type.
Appearance	It usually appears in the raceways, as the free water in the oil tends to sink (the first is heavier than the second), creating small gap between the roller and the raceway.

Frictional corrosion - fretting corrosion

Aspect	The bearing presents corroded areas and fracture notches.

Causes — It is caused by an inappropriate fit between the bearing ring and the shaft/-housing or by the existence of inaccuracies. The relative movement detaches small particles from the surface that oxidise quickly in the existing air.

Acceptability — The bearing remains in an acceptable condition while the surface is worn away to a depth smaller than 0.100 mm. In some applications, if the depth exceeds this value, the bearing can still be used if the fretting marks are removed by grinding.

Solutions — Apply anti-fretting paste on the surface, use polyamide spacers (between the backing ring and the inner ring).

Appearance — It appears on the bore of the inner ring or in the outside of the outer ring.

Frictional corrosion - false brinelling

Aspect — There are shallow depressions formed on the raceway and the rolling elements are often discoloured or shiny.

Causes — It occurs due to the contact of two surfaces due to cyclic vibration - micro-movement or resilience of the elastic contact.

Acceptability — If the damaged can't be felt by a finger nail or of it can be removed by polishing with an abrasive paper, the bearing is still acceptable.

Solutions — Reduce the existing vibration by applying an anti-vibrational solution.

Appearance — It appears on the raceways and rolling elements.

2.2.4 Electrical Corrosion

It is caused by changes in the material structure due to a passage of an electric current.

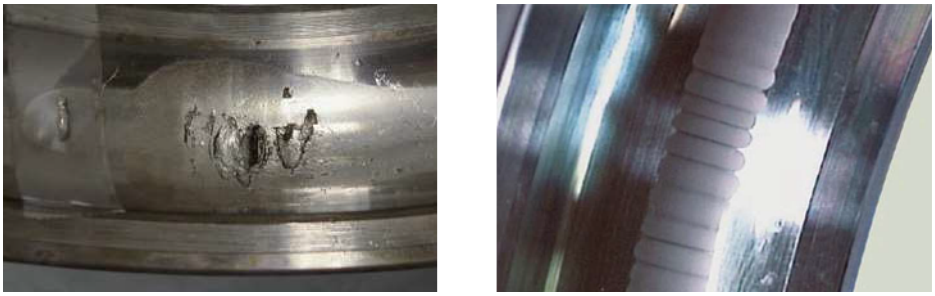


Figure 2.12: Example of excessive voltage on a ball bearing raceway (on the left) and washboarding in a raceway of a deep-groove ball bearing (on the right) [6].

Excessive voltage

Aspect — The components presents discoloured areas where the material as been tempered, re-hardened or, when there exist craters, melted.

Causes — It is caused by the passage of an electric current from one of the rings to the other via the rolling elements, causing a similar effect to and electric arc welding, as there is a hig current density over a small contact surface.

Acceptability A bearings presenting craters shall be considered scrap.

Solutions Use an earth connection or an earth return device.

Appearance It appears as craters or localized burns in raceways and rollers or as zigzag burns in the raceways.

Current leakage

Aspect The components presents mini-craters closely positioned, of minor size when compared to the ones formed by excessive voltage. Sometimes, washboarding (or fluting) is developed from the craters.

Causes Is occurs due to a current flow, even at low intensity, that heats and erodes the components.

Acceptability A bearings presenting this kind of damage shall be considered scrap.

Solutions Use an earthing device.

2.2.5 Plastic Deformation

It is caused by excessive and concentrated forces and stress applied that deform the bearing.

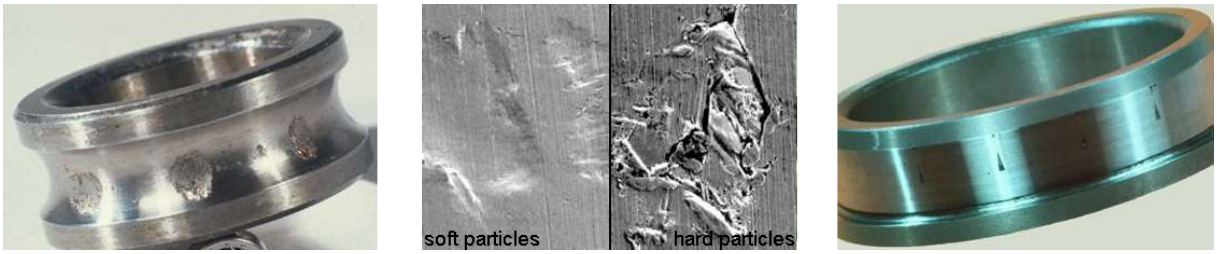


Figure 2.13: Example of overload from wrong mounting (on the left), identation from debris of soft and hard particles (on the center) and indentation from handling on a inner ring of a cylindrical roller (on the right) [6].

Overload

Aspect The components show some considerable plastic deformation.

Causes It is typically caused by incorrect mounting or blows in the components but can produced by other factors, as excessive and concentrated stress or excessive forces.

Acceptability A bearing showing this kind of damage in no longer acceptable.

Solutions Use adequate mounting tools.

Indentation from debris

Aspect	The bearing shows heavy indentations and raised material.
Causes	It is caused be foreign particles that entry the interior of the bearing or by contaminants existent in the lubricant. They produce indentations that can lead to plastic deformation, fatigue and premature spalling
Acceptability	The bearing can be acceptable if the damage is presented only slightly and not across the entire raceway.
Solutions	Replace seals, replace the grease with a clean one.
Appearance	It appears usually in the raceway.

Indentation from handling

Aspect	The components present nicks, caused be localized overloading.
Causes	It is caused by inadequate handling during transport, stoling, mounting.
Acceptability	If the damaged occurs in an isolated area and isn't deeper than 0.050 mm the bearing can still be used.
Solutions	Handle the bearing with care.

2.2.6 Fracture and cracking

It is characterized by the break and loss of material caused by abnormal forces and stress applied in the bearing.

Forced fracture

Aspect	The bearings exhibits fractures.
Causes	Is is caused by stress concentration that exceeds the tensile strength, created by local overloading or over-stressing.
Acceptability	An bearing showing this kind of damage in no longer acceptable.
Solutions	Use the correct mounting tools, check if the bearing as the correct size.

Fatigue fracture

Aspect	All the component is cracked, although in an early stage it only shows initiated cracks.
Causes	It occurs due to load cycles in which the fatigue strength is exceeded.
Acceptability	Any bearing showing this kind of error should be scrapped.
Solutions	Use a correct mounting procedure, making sure that the bearing seats are correct.

Thermal cracking

Aspect	The components exhibits cracks, usually perpendicular to the direction of movement.
Causes	It happens due to an incorrect mounting, as it allows the relative movement between components that should be locked, causing a high frictional heat.
Acceptability	Any bearing showing this kind of damage should be considered scrap.
Solutions	Follow a correct mounting procedure.

Figure 2.14: Example of forced fracture on a bearings outer ring provoked by a bad housing seat (on the left), fatigue fracture of a spherical roller bearing outer ring (on the center) and thermal cracking resulting on transverse cracks on the side face of a bearing inner ring (on the right) [6].

2.3 Bearings characteristic frequencies

When a bearing presents defects, it behaves differently from a bearing in a good condition. When a defect in the surface of a component strikes another surface, an impulse is generated (when the rolling elements strikes a local fault on the inner or in the outer race, or when a fault on a rolling element strikes either the inner or outer race). The rotation of the faulty component will produce a series of successive impulses [8]. The impulse characteristics will depend upon the operating speed, geometry of the bearing and decay in vibration transmission path (dependent of the system damping factor). At a constant rotating speed, the signal will present peaks produced at a certain frequency called bearing characteristic frequency (the signal may be considered cyclostationary). Other peaks may be present, produced by the excitation of the bearing system resonance [9].

Depending on the location of the fault (inner race, outer race or roller), three different frequencies may be produced. It may be of interest to calculate also the frequency related to the movement of the cage. Bellow, the formulae of the various frequencies will be displayed [9]:

Ball pass frequency, outer race:

$$\mathrm{BPFO} = \frac{nf_r}{2}\left\{1 - \frac{d}{D}\cos\phi\right\} \tag{2.1}$$

Ball pass frequency, inner race:

$$\mathrm{BPFI} = \frac{nf_r}{2}\left\{1 + \frac{d}{D}\cos\phi\right\} \tag{2.2}$$

Ball spin frequency:

$$\mathrm{BSF} = \frac{D}{2d}\left\{1 - \left(\frac{d}{D}\cos\phi\right)^2\right\} \tag{2.3}$$

Fundamental train frequency (cage speed):

$$\text{FTF} = \frac{f_r}{2}\left\{1 - \frac{d}{D}\cos\phi\right\} \tag{2.4}$$

where d is the roller diameter, D the pitch diameter (diameter of the circle containing the center of all rollers), f_r is the shaft speed (in Hz), n is the number of rolling elements (per row) and ϕ is the angle of contact. It should be noted that the ball spin frequency (BSF) is the frequency in which the fault strikes the same race (inner or outer). If the frequency of shocks produced by the roller is to be known, the value considered should be the BSF multiplied by two.

The frequencies calculated through this formulae will present some deviation from the real value, typically in the order of 1-2% due to two different factors: random variation around the mean frequency, and because of the slip that must virtually exist (the kinematic frequencies were calculated assuming no slip). Each rolling element will have a different contact angle (ϕ) as the ratio of local radial to axial load will change at each position. This will force each rolling element to have a different effective rolling diameter, trying to roll at a different rolling speed at each position. The cage will limit the deviation from their mean position, causing some random slip. While small, the random slip will change the character of the signal [9].

Although the formulation of the frequencies presented above refers to ball bearings, it can still be utilized for roller bearings. In this case, the d shall refer to the average diameter of rollers, and for tapered bearing the average contact angle shall be considered for the value of ϕ. Note that the argument about the variation of rolling diameter applies equally to tapered and spherical roller bearings, but shouldn't be applied to parallel roller bearings (in this case an angle of 0 should be considered).

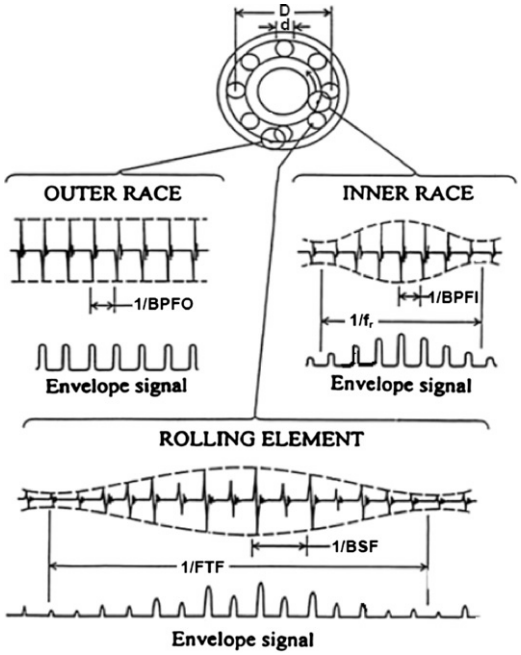


Figure 2.15: Typical vibration signals and envelope signals from local faults in rolling ellement bearings [3].

The series of broadband bursts excited by the shocks are modulated in amplitude, by two factors [9]:

- The strength of the bursts is normally modulated by the rate at which the fault is passing through the load zone (the strength of the burst depends on the load borne by the rolling element).

- The transfer function of the transmission path varies with respect to the fixed positions of response transducers, for cases in which the fault is moving.

The figure 2.15 illustrates typical modulation patterns, considering unidirectional (vertical) load on the bearings. With this figure is clear that the analysis of the envelope signal is a valuable tool, that gives more diagnostic information than the analysis of the raw signal.

Analysing the bearing vibrational signal at these characteristic defect frequencies can be a valuable procedure to detec a defect, estimate its size and determine its location [10]. Its is important to estimate beforehand the various frequencies to link, depending on the frequency of the burst and its harmonics, the detected fault to its location. Through its amplitude, the severity of the fault can be estimated.

Chapter 3

Railway

The technology of the railway systems has suffered different evolution phases through history. In the late 18th and the early 19th, the strength of the materials used proved to be a limiting factor, preventing the vehicles of reaching high velocities in which the contribution of the dynamics effects was relatively small. The use of a flanged wheel was one of the earliest innovations and solved the major existing problems as, for example, the guidance of the train through the tracks.

With the discovery of new material, the application of new mechanical concepts and the appearance of the electric locomotive, the size of the railway vehicle and its maximum velocity increased. The dynamic effects started to be accounted and expressed in mathematical formulations in the late 19th. Solutions for the oscillation of the car and the forces generated in the negotiation of the curves were widely pursued, aiming for the improvement of the ride quality and safety.

Through the 20th century, the yet increasing speed of trains and the greater potential risk arising from the existing instabilities obligated new scientific approaches. Complex mathematical approaches could only be solved with the appearance of the power of digital calculation in the last half of the 20th century.

One of the most important structural changes in early models was the introduction of bogies in 1840. Today, there isn't a single train operating without this component. As it comprises the wheelset and axle box, where the bearing is located, we will pay closer attention to this structure.[11]

3.1 Bogies

Originally the first railway vehicles, two-axle vehicles, had a pair of wheelset (the wheel and axle assembly) incorporated in the carriage. The rigid wheelbase introduced various problems to the main movement of the train in the curves, as in the ideal scenario each wheelset axle should be aligned with radius of the curvature. In this configuration, they were always aligned in a perpendicularly direction of the movement of the car. This imposed various constrictions, as the necessity of smooth and longer rail-track curves and shorter carriages.

The solution found for this problem was the transference of the wheelset to a external structure - called bogie - establishing the connection between the carriage and the track. It not only suppress the problem of the circumscription of the carriage in the curves, but allows some other improvements as the integration of a braking system and the main and secondary suspension.[11][12]

Its main functions are the following:

- Ensure the connection of the carriage to the bogie, through central pivot;
- Support the railcar body firmly;

- Incorporate the suspension and bearing system;
- Ensure the correct movement of the overall vehicle and its stability in the straight and curved tracks;
- Increase the comfort by absorving vibration, minimizing centrifugal forces on curves at high speed;
- Minimize generation of track irregularities and rail abrasion.

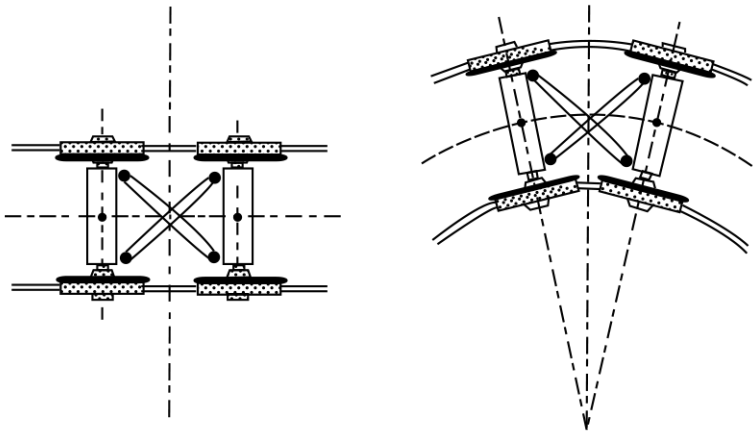


Figure 3.1: Circumscription of the carriage with the track, achieved through the use of a bogie [11].

3.1.1 Types of bogies

The bogie can be classified into different types according to its different characteristics. [12]

Number of axles It can be classified as *single-axle*, *two-axle*, *three-axle*, and so forth.
The single axle type is the lowest classification possible and the two-axle the most common type. The two-axle type is the simplest solution that assures a good performance, as it reduces the impact of the track irregularities, unlike the single-axle type that transmits them directly to the car suspension point. The three-axle bogie as a worst running performance mainly because of it's complex structure.

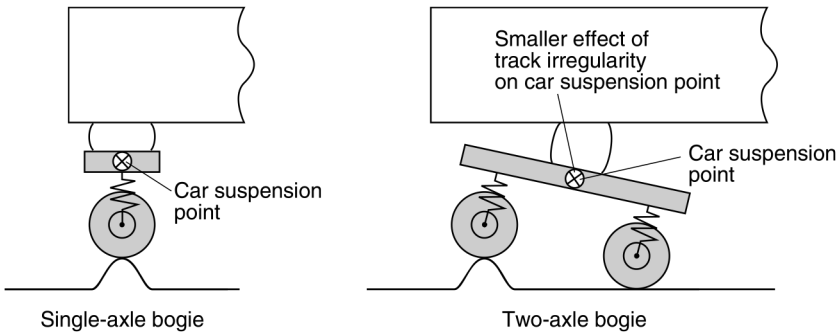


Figure 3.2: Single-axle and two-axle bogie and the comparison of of effect of track irregularity [12].

Existence of an articulation It can be classified as an *articulated* and *non-articulated type*.
The articulated type supports the back end of the forward car and the front end of the rear car while the non-articulated type is applied in sets of two in each car. The articulated type offers a lower gravity center, better ride confort and less effect on running noise, as the bogie is not overhanged by the car and, consequently, isn't placed directly under the seats of the passengers. It is a more complex solution of a more difficult maintenance and increases the axle load, as each body is supported by one bogie.

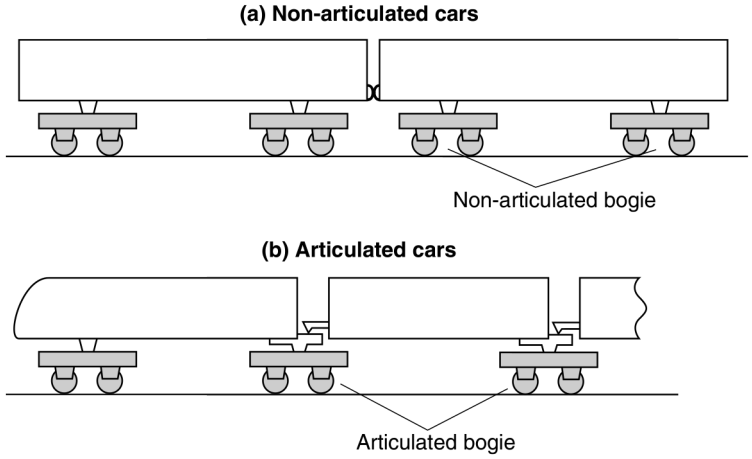


Figure 3.3: Articulated and non-articulated bogie [12].

Suspension gear It can be classified as *bolster* or *bolsterless*.
Although the bogie must rotate away from the body on curves, it is of prime importance that it secures high rotational resistance in straight sections to prevent wheelset hunting, which reduces the ride comfort and may lead to derailment. The first solution found, classified as bolster bogie, was the use of side bearers that resists rotation with the central pivot. In 1990 a new solution started being commercialized, the bolsterless bogie, that uses a bolster springs that deform on the curves, thus permitting rotational movement, along with longitudinal anti-yawning damper, that restrains the rotational movement in the straight sections. This new solution proved to achieve better results while reducing the number of parts used and the body weight.

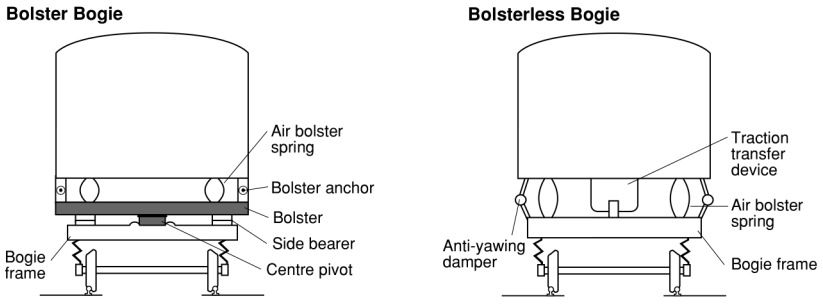


Figure 3.4: Bolster and bolsterless bogie [12].

Suspensions structure It can be classified as *swing hanger* or *small lateral stiffness bolster spring*.
The bogie must absorb rolling motion in order to assure a good quality ride to the passengers. There are two possible systems that satisfies this purpose, the swing hanger bogie that supports the body by using a lower swing bolster beam suspended from the bogie

frame by means of a link consisting of two bottom-widening vertical members, bolster springs and upper swing bolster beams, and the small lateral stiffness bolster spring bogie, that supports the body by using bolster with air bolster springs and bolster anchors. The latter solution enables a reduction on the weight and size of the overall structure and a easier maintenance due to the simplifies suspension design, making it the preferred and must used solution nowadays.

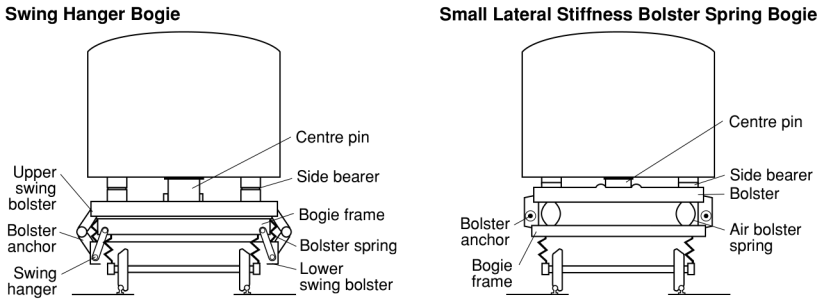


Figure 3.5: Swing hanger and small lateral stiffness bolster spring bogie [12].

3.1.2 Bogies components

The bogie may be composed by a various range of components depending on the type of the bogie.

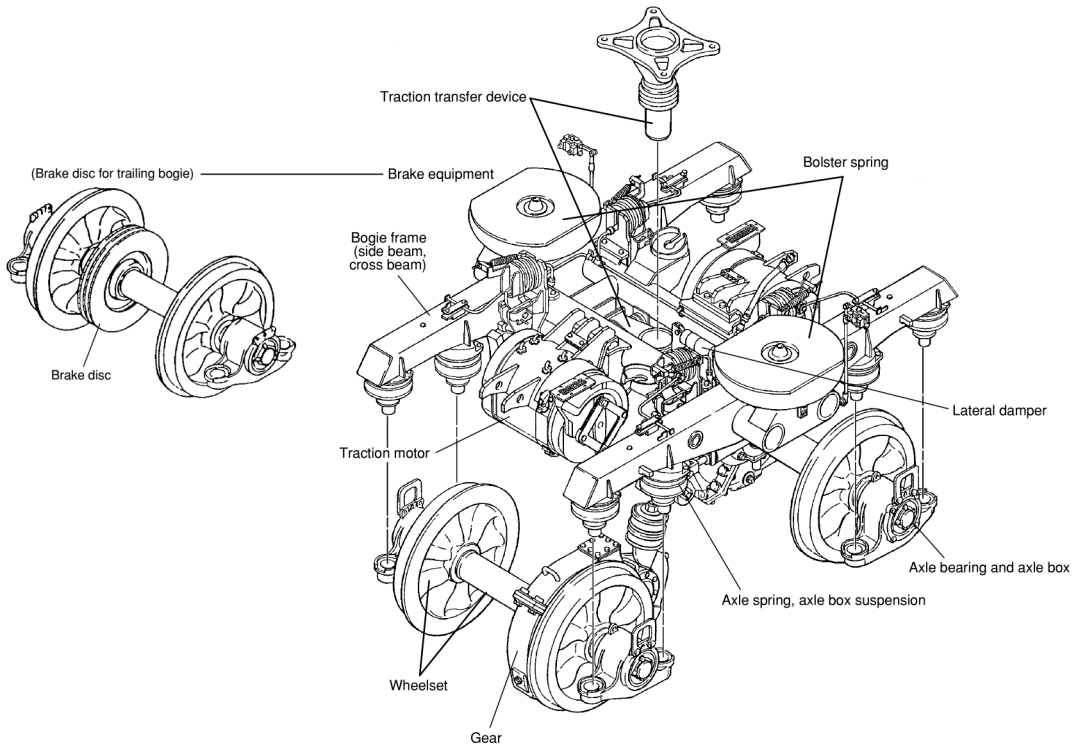


Figure 3.6: Components of a bogie (DT50 Bogie model) [12].

The following components are found in every existing bogie and must exist to ensure its right behaviour[12].

Suspension gear Its main function is to support the body of the train, allowing the bogie to rotate relative to the car in the curves and transmit traction force from the bogie to the car. In most cases it is composed by bolster springs, traction transfer device, anti-yawning damper and lateral damper.

Bogie frame It is the components that accommodates the various bogie equipment and in most cases its a welded construction.

Axle box suspension This component assures a good performance of the bogie and the ride comfort. It supports the axle via the bearing from the bogie frame and there is a wide range of designs used for this purpose.

Wheels This component promotes the movement of the train by rolling along the tracks. Is a conic profile and a flange that prevents it from running of the track. At certain velocities the wheels are prone to mass imbalance, producing resonance vibrations.

Axle The axle secures the connection of the wheels to remaining equipment. It can be solid or hollow and is supported by the bearings.

Axle bearings The type of axle bearings differs, but it is usually used tapered rolling bearing or a combination of cylindrical roller bearings and ball bearings

Transmission It transmits the motive power generated by the motor or engine to the axle. It is composed by gears and a flexible coupling.

Brake equipment The brake equipment permits the vehicle to stop at a desired position. There most common mechanical equipments encountered in the brakes are wheel tread brakes and disk brakes.

3.2 Axle-bearing for railway applications

As it was already mentioned in the last chapter, the rolling element bearings are a fundamental part of a railway vehicle structure. There are various sub-types that can be used, each one having its advantages and disadvantages.

Although in an early period it was tried the use of plain bearings in this equipment, it proved to be a poor choice as they provided a high friction coefficient when starting from rest, a poor reliability and labour-intensive maintenance. However, there were some recent attempts to reintroduce bearings of this type that did not require lubrication, providing some positive features from a vehicle dynamic behaviour point of view.

In the railway industry, we can find the following type of bearings in the axle boxes: tapered roller bearings, cylindrical roller bearings and spherical roller bearing.
Spherical roller bearings are rarely used although they provide a better distribution of load in case of axle bending in which axial forces are transmitted. As they have a high cost and a lower weight capacity, when utilized they are combined with cylindrical bearings.
The *cylindrical roller bearings* main disadvantage is its inability to transmit axial forces but they have a high capacity in the radial direction. Regulating the diameter, length of rollers, radial and axial clearance it is possible to reach a state in which the faces of the rollers can resist lateral forces. In high speed rolling stocks, they are often combined with *ball bearings*, that supports the axial force.
The *tapered roller bearings* can transmit axial and radial forces due to its inclination to the rotation axis, although for an optimal response the tolerances of the rollers and their clearance must be kept at tighter levels than the one used for cylindrical bearings[11].

3.3 Types of axle failures

A bearing is specially chosen to operate at certain specifications. As in the railway application the lifetime of a bearing never exceeds the vehicle's one, the bearings must be replaced when it reaches a acceptable limit. The smallest unpredictable alteration may provoke damages on the bearing that will shorten its life. Being so, it is not only important to study the failure of the bearing but other external failures.

The classification system (axle failures) that is displayed below may rise some doubts about its connection to the one presented in the previous chapter (rolling bearing failure). The first classification system that was presented is a more general one. It was developed by bearing constructors and specialists to analyse the damage progression on bearings. The second classification system was developed to classify failures in the railway areas. It is of easier application, as it focus not only on bearing's observable defects, but can be directly deduced through signs easily observed in the vehicle. However, it may be important to bear the first classification system in mind, as it offers a more theoretical perception of the damage.

Table 3.1 offers a direct relation between the axle failures and the bearings failures.

Based on data from the Maxbe Project [13], we can acknowledge the most frequent axle-failures in some railway operators. The percentage displayed below correspond to the frequency of occurrence of each failure.

- **Loose bearing failure** 23%;
- **Water etch** 20%;
- **Wheel tread defect** 13%;
- **Fatigue spalling** 13%;
- **Bearing destruction** 12%;
- **Mechanical** 6%;
- **Lubrication** 4%;
- **Adapter** (Displaced, worn, wrong size or broken) 4%;
- **Displaced seal** 3%;
- **Truck related failure** 2%;
- **Application defects** 0%;
- **Manufacturer, remanufacturer or reconditioner defect** 0%;

Loose Bearing Failure

Symptoms Usually, the main symptom is the presence of a small gap (grater than $0.070''$) between the bearing inner diameter and the axle journal. Other symptoms can be present, as loose cap screws, loose backing ring, spun cone, cone back face wear, etc.

Causes It can start due to two different groups of causes: a oversize cone bore or undersized journal; or due to low initial clamp load, low interference fit or heavy load for a long time. This, in conjunction with a loss of lateral clamp load (originated by axle flexure) will lead to loss of cone fit. Eventually, it leads to point loading in the rollers and cupe race small end. As fatigue increases, the rollers will skew, leading to obstruction and the destruction of the spun cone as it takes full load.

Water etch, displaced seal, fragment indentation and wheel defects with no broken cage may be observable, but the overall evidence leads to this type of failure.

Figure 3.7: Damage on a bearing that heated to unacceptable temperatures due to loose bearing components [13].

Water Etch

Symptoms Usually, the main symptom is etching of races and rollers, occurring at a depth larger than 0.010" and at irregular intervals less than a roller width apart. Other symptoms can be present, as rusty components, water each spalling, water or moisture in grease, severe pitting on race edges, etc. Staining and light etch signs that could have occurred after service shouldn't be considered.

Causes It starts with the ingress of moisture or with contamination of lubricant. It eventually leads to race and rollers etching, proceeded by spalling. The wear of the cage (if no obstruction and destruction of the bearing occurs in the meanwhile) will first lead to the skewing the rollers, latter to the smearing of the race and finally to the production of heat.

Loose stack, brinelling of raceways, and broken cage (with the wheel in a good condition) may be observable, but if the overall evidence match the ones mentioned above, water etch failure should be considered.

Figure 3.8: Sign of water ingress (on the left and center) and raceway etching (on the right) are signs of water etch failure [13].

Wheel Tread Defect

Symptoms Out of round (one depression deeper than 0.070″ or various deeper than 0.05″). or wheel impact detection superior to 90 kips and broken cage. Other symptoms can be present as shelling, tread built up, out-of-round or slid flat, skewed rollers, smearing races and rollers and transference of material between the bearing's components.

Causes It usually starts with a defect in the wheel tread. It leads to impacts and the cage to broke. The rollers will skew, provoking the smearing and transfer of material of races and rollers, heating the bearing to unacceptable temperatures.

Other symptoms as water etch, light brinelling, back face wear or loose cap screws may be present, but if the overall evidence match the ones mentioned above, wheel tread defect failure should be considered.

Figure 3.9: Whell tread defects (on the left) usually leads to bearaing failure as, for example, breaking of bearings cage (on the center and right) [13].

Fatigue Spalling

Symptoms Usually, the main symptom are signs of spalling over the races (geometry problems, water etch, impact indentation or signs of an incorrect montage of external components shouldn't be present). It often occurs in repaired raceways - which have been reground or in which the spalls have been repaired - that show fatigue.

Causes The spalls are originated by cyclical fatigue. It occurs randomly through the raceways and rollers surface in an early stage, and tend to appear near the center of the components in an advance state. As the spalls propagate, there is smearing of raceway and roller and heat is produced.

Other defects can be present, as water etch, impact brinelling, uneven load, defective lubrication but they never are the cause for spalling. Fatigue spalling may be originated from steel defects, reground or repaired raceways or severe loading.

Figure 3.10: Examples of bearing distress caused by fatigue spalling [13].

Bearing destruction

Symptoms This classification is used in cases where the cause of failure isn't clear. The cone of the bearing tend to be broken and fused to the journal.

Causes An unknown cause provokes the heating of the bearing, which eventually contributes to the malfunction of the components that are destroyed or fused to others.

Some problems are linked to this type of failure. The presence of loose components (as the cone, wear ring, backing ring, cap screws and even the bearing), water etch, wheel defects or other mechanical problems can assure this type of failure.

Figure 3.11: Examples of bearing distress with unkown causes [13].

Mechanical

Symptoms Mechanical problems, as brinelling on the raceway, external heat evidences, cup fracture, fluting or electrical arcing, substances wrapped around or pulled into the bearing, damaged seals are some of the symptoms of this kind of failure.

Causes The origin of the problem is usually due to physical abuse of the bearing, that lead to the damage of the seal an loss of lubricant, leading to heat generation and brinelling of raceways. The brinelling usually leads to small spall zones that propagate through all the bearing.

Figure 3.12: Signs of external abuse that lead to bearing distress [13].

Lubrication

Symptoms The symptoms can be directly link to the lubricant problems, as loss of grease, thin/thick and runny grease or burnt smell, or indirectly, as blue races and roller, possibly showing signs of scoring, smearing and peeling, polished raceways and spalling of cone.

Causes It usually starts with a modification in the quantity or consistency of lubricant. The rollers end start to appear scored, and the raceways polished. The components start to smear and spall, heat is produced, and eventually the cage wears (or even breaks), skewing the bearings.

Any of the previous causes can produce enough heat to damage permanently the bearing. Although spalling can produce residues that will contaminate the lubricant, isn't associated with this failure mode.

Figure 3.13: Bearing distress due to lubrification break down [13].

Adapter (displaced, worn, wrong size or broken)

Symptoms A displaced, worn, wrong sized or broken adapter can lead to various problems as wear of the end cap, back ring or seal case, point loading in or cracking of the cup, uneven loading and the wear and broking of the adapter itself.

Causes An displaced adapter will lead to point loading on races, provoking spalling and smearing of races and rollers. It can displace the seals, leading to loss of lubricant. It can displace adapters, leading to rubbing on several parts of the bearing. All of these causes the heating of the bearing and its malfunction.

Displaced seal, race spalling and grease contamination may be observable, but they represent the consequences of this failure mode and not the failure mode itself.

Displaced seal

Symptoms The displacement of the seal and its rubbing against other components, presence of free water in the bearing, loss of lubricant and uneven wear of the seal.

Causes The displacement of the seal lead to its rubbing against other components, permitting the loss of lubricant and entrance of water. This can lead to smearing and spalling of the races and rollers, producing excessive heat.

This failure mode should be attributed to cases where the displacement reason isn't evident. Possible causes are improper mounting loose bearing components and debris or contact between seal and adapter.

Truck related failure

Symptoms Usually, the main symptom is heavy worn of bearing components, spalling in raceways, with a failure date inferior to 4 years since the mounting date.

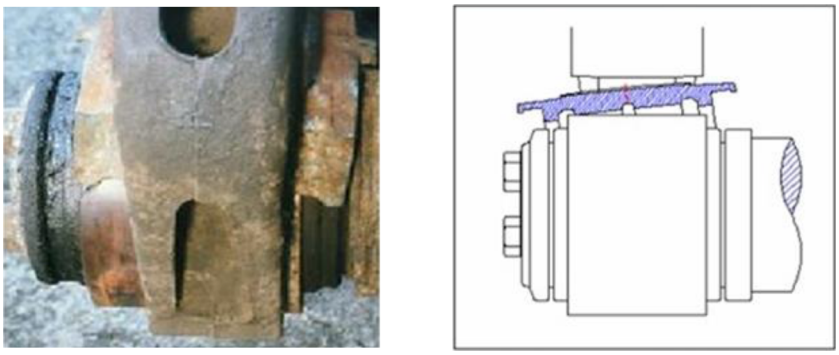


Figure 3.14: Bearing distress (on the left) caused by displacement of the adapter, which is represented on the figure on the right [13].

Figure 3.15: Displaced seals and consequential bearing distress [13].

Causes The damage starts due to an abnormal performance of the truck. The cone race, and then the cup race and rollers start to wear, to spall and to smear. The load is transferred to the outboard race, provoking the truck to skew in curves. Then, inboard and outboard race limits start to suffer from severe spalling and wear from overload, until all the cup undergoes fatigue wear.

As the cause of this failure is due to the car assembly's abnormal performance, the failure will tend to repeat itself along the truck's life.

Application defects

Symptoms There are various possible symptoms, as the presence of wrong sized or out of tolerance components, misplaced or missing elements, observable handling and mounting damage as stress markings, sheared or loose components.

Causes Several causes can lead to bearing failure:
Misalignment of the mounting press is often the principal cause for this type of failure. It can lead to gouging of bearing components or seal damaging.
Undersized journals will permit the cone to spin, loosening the bearing
Components are dropped in the mounting process and are sheared.

Water ingress or journal grooving may be observable, but the overall evidence leads to this type of failure, although it should only be considered if the bearing is less than 4 years old since mounting date.

Manufacturer/remanufacturer/reconditioner defect

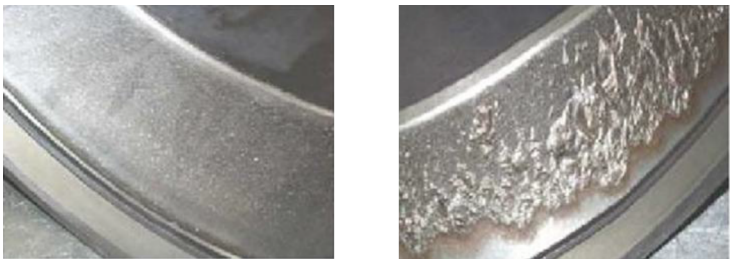


Figure 3.16: Bearing distress caused by uneven loading from truck components [13].

Figure 3.17: An incorrect assembly procedure (on the left) can cause bearing distress (on the right)[13].

Symptoms Presence of mixed components, presence of components with wrong dimensions, bearing assembly defects, wrong grease charge, or presence of defective manufacturer components or malfunctioning remanufactured/reconditioned components.

Causes The assembly defects will evolve and end up causing the premature failure of the bearing.

Although the symptoms are similar to the ones that cause Application Defect failure, this present classification is attributed to improper manufacturing, remanufacturing or reconditioning of the bearing. This type of failure will tend to repeat itself among several products supplies.

Figure 3.18: An improper manufactiring (bearing dimensions being checked on the left image) can cause bearing distress (signs of etch on the rollers, on the left and middle images) [13].

It is important to connect the two classification systems as to allow a more theoretical study on railway vehicle's failure. It should be noted that other bearing failures may be present, but the ones indicated below are the most probable.

Table 3.1: Axle failure and the corresponding bearing failures.

Axle failure	Related Bearing Failure		Notes
	Early stage	Advance stage	
Loose bearing failure	Sub-surface initiated fatigue	Forced fracture	Usually, the first failure felt on the bearing is sub-surface initiated fatigue due to the bearing being loose. In latter stages, the obstruction of the bearing will lead to forced fracture.
Water etch	Moisture corrosion and Surface initiated fatigue	Adhesive wear and Forced fracture	In earlier stages it is linked to moisture corrosion (due to the ingress of moisture), that will evolve to surface initiated fatigue as the bearing spalls. If the bearing is not repaired/replaced, it will eventually lead to adhesive wear and forced fracture.
Water etch	Moisture corrosion and Surface initiated fatigue	Adhesive wear and Forced fracture	In earlier stages it is linked to moisture corrosion (due to the ingress of moisture), that will evolve to surface initiated fatigue as the bearing spalls. If the bearing is not repaired/replaced, it will eventually lead to adhesive wear and forced fracture.
Wheel tread defect	Forced fracture	Adhesive wear and Abrasive wear	A external defect, in the wheel thread, affects the bearing, usually leading to forced fracture in some components. As it worsens, other failures as adhesive and abrasive wear may start to progress.
Fatigue spalling	Sub-surface initiated fatigue and Surface initiated fatigue	Adhesive wear	The cyclical fatigue usually appears in the form of sub-surface initiated fatigue and, after a while, can lead to surface initiated fatigue. In an advanced state, signs of adhesive wear will start to appear.
Bearing destruction	Unknown	Forced fracture, Overload and Thermal cracking	The cause to this type of failure isn't clear. As it can be related to other axle failures as loose bearing faults, water etch, wheel tread defect and mechanical problem, the respective bearing failure defects should be considered. Usually, bearing destruction failure is only detected in advanced stages, where forced fracture, overload and thermal cracking are present.
Mechanical	Indentation from debris and surface initiated fatigue	Abrasive wear	The physical abuse of the bearing usually manifests itself as indentation from debris, that may lead to surface initiated fatigue. Signs of abrasive wear may be present at advanced stages.

Table 3.1: (continued)

Axle failure	Early stage	Advance stage	Notes
Lubrication	Abrasive wear, Adhesive wear and Surface initiated fatigue	Overload and Forced fracture	Lubricant-related problems are usually connected to abrasive wear, but may be related also to adhesive wear. In time, it can lead to surface initiated fatigue. In latter stages, it can provoke overload and forced fracture.
Adapter (Displaced, worn, wrong size or broken)	Adhesive wear and Sub-surface initiated fatigue	Abrasive wear and Fatigue fracture	Problems in the adapter can lead to adhesive wear or to sub-surface initiated fatigue. It may evolve to abrasive wear, and then to fatigue fracture.
Displaced seal	Moisture corrosion, adhesive wear and sub-surface initiated fatigue	Abrasive wear	The displaced seal can allow the entrance of water that leads, in time, to moisture corrosion, but can also contribute to adhesive wear and sub-surface initiated fatigue. In advanced stages it can lead to abrasive wear.
Truck related failure	Adhesive wear and Sub-surface initiated fatigue	Surface initiated fatigue and Overload	An abnormal performance of the truck may lead to adhesive wear or sub-surface initiated fatigue. In latter stages, signs of surface initiated fatigue and overload are present.
Application defects	Overload, Indentation from handling and Forced fracture	Various (related to fatigue, wear, fracture and cracking)	An incorrect mounting procedure produces signs of overload, indentation from handling and forced fracture. The components of the bearing present this kind of damage as soon as they are applied. The bearing should no operate under this conditions, but if it does, its bad performance can lead, in more advanced stages, to various kinds of damage related to fatigue, wear and fracture and cracking.
Manufacturer, remanufacturer or reconditioner defect	Overload, Indentation from handling and Forced fracture	Various (related to fatigue, wear, fracture and cracking)	This problem evolves to the use of incorrect components. Causes related to this kind of fault usually manifests soon and have a fast degradation speed, being detected after a short period of time. The bearing failures related to this type of damage are the same as the types of damage related to application defects.

3.4 Asset management in railway organizations

Asset management is a term that was initially associated with the financial sector but has lately been applied to diverse organisations, mainly the ones responsible for infrastructure networks in which the cost and performance of their infrastructures are of national importance. It is defined in by the British Standards Institute Publicly Available Specification for the management of physical assets (BSI PAS 55-1) as the *Systematic and coordinated activities and practice through which an organization optimally manages its assets and their associated performance,*

risks and expenditures over their lifecycle for the purpose of delivering the organization's business objectives[14], in which the word asset should be understood as *Plant, machinery, property, buildings, vehicles and other items that have a distinct value to the organisation*[15].
The main objectives of asset management can be defined as follows[14]:

- Establish a organisations' strategic plan, that pays special attention to safety, performance, capacity, availability and sustainable development;
- Understand the required asset performance, condition, cost and risk;
- Use whole-life management tools and techniques to seek cost-benefit optimisation;
- Prioritise capital investment, based upon the assets contribution to service.
- Monitor, forecast and improve performance of the assets.

The railway plays a crucial role in stimulating economic growth, as there is a ever demanding necessity for safer and quicker rail transportation. In the railway infrastructure, asset management is about delivering, in a safe, reliable and sustainable way, the outputs valued by customers, funders and other key stakeholders for the lowest whole life cost. As the railway sector can be categorised as a high-hazard sector, additional safety requirement are needed, requiring more precise planning and safeguards design for each discrete activity. It should be assured that the right decisions to build safety into the physical environment and into the way people (workforce, passengers and public) interact with it are made.
It is applied to various activities of a railway organization, as demand forecasting, strategic planning, investment decision making, annual activity planning, work scheduling, design development, project and work delivery, operation, incident management, assurance, review and learning.

Diverse paths can be drawn, but they all will eventually cover the maintenance, renewal and enhancement in activities that will delivery sustainable outputs at the whole-life cost (opposed to prioritising work predominantly according to condition or reliability alone). The optimization process should bear in mind the following subjects:

- The safety risk to passengers, workforce and members of the public;
- Impact of infrastructures faults on train performance;
- Impact of the infrastructure on the environment;
- Life of the infrastructure;
- Lifecycle cost requirements and assessments of asset residual life;
- Resilience of the infrastructure to weather and climate change.

All of the decisions should be evidence-based, using the knowledge of how assets both degrade and fail, in order to optimise the interventions.

The Asset Management Consulting Ltd (AMCL Ltd) designed an asset managment model consisting of six main groups: asset management strategy and planning; asset management decision-making; lifecycle delivery activities; asset knowledge enables; organisation and people enables; risk and review. Details of each group subjects can be found in figure 3.19.

The proper application of asset management can bring positive results. For example, in 2009 the Network Rail reduced its maintaining, renewing and operating cost by 27%, despite the increase of 5% in the number of train kilometres covered.

Figure 3.19: Asset Management Consulting Ltd's asset management model[15].

3.4.1 Asset polices

Asset policies play an important role in asset management. It represent the *Specified approach, rules and boundaries that provide direction and the framework for the control of specific asset-related processes and activities such as capital investment and maintenance*, according to BSI PAS 55:2008.
Asset polices should be aligned with the asset management objectives. They should be establish based on the minimization of whole life and whole system costs, and supported by facts from the asset information. The most appropriated approaches to asset maintenance, inspection and renewal must be designed focusing on reliability, availability, maintainability and safety specification[15].

There is a large scope of assets groups in an railway organization: ground area, track and track bed, engineering structures (bridges, culverts, overpasses, tunnels, etc), level crossing, superstructures (rails, sleepers, ballast, points, crossings), access way, safety and signalling installations, telecommunications and lighting installations and railway vehicles. As all the assets are eventually interrelated, decisions should be well considered, as the activity on one asset group may impact another part of the overall railway system.

With each asset there is a range of decisions to be made (on inspection, maintenance, renewal and enhancements). So, a strategic that covers the following aspects must be drawn[16]:

- Inspection regime (type of monitoring technique);
- Maintenance and renewal intervention criteria (when to intervene and how to intervene);
- Equipment obsolescence assumptions;
- Technical strategy
- Redudancy
- Cost of inspection, maintenance and renewal.

The information collected can be related to asset type/location, age, capability, condition, failure history and consequence, work histories, unit cost and as-built drawings. The existence

of asset's information is critical to the proper planning of maintenance and renewal actions for each type of asset. It can be achieved through enhanced mechanisms for automated data capture that capture and collate and the data.

One indispensable tool for the understanding of the asset performance and of rates of asset degradation is condition monitoring technology. It is usually applied to critical assets as the cost related to the implementation of condition monitoring limits the range of assets in which condition monitoring is profitable.
It helps to intervene before individual assets fail and, if used with a proper risk-based maintenance program, condition monitoring enables to maintain the infrastructures in a reliable way at a lower cost.

Risk-based maintenance proved to be the best technique to reduce the total cost of the whole-life equipment cost. After the establishment of the life cost model and the possible failures modes, methods that evaluate the acceptability of the consequence of failure are used to establish maintenance regimes. It enables the construction of a system that tests the possible trade-offs: different levels of risk are defined regarding each possible maintenance or renewal action. Having study how the assets degradation and failure impacts the train service and the safety of passengers, workers and other members, it is then possible to find specific cases where it better to do more maintenance, extending the assets life, or the opposite. The reduction of maintenance may have better consequences as the actual implication of not carrying the maintenance is less costly than carrying out the task (proving that the undue risk is not a problem).

3.4.2 Condition monitoring in railway organisations

As already seen, condition monitoring proves to be a tool used for proper asset management. A more general approach will be made in the following chapter. In the present section, we will synthesize the advantages of the application of condition monitoring in railway vehicles.

The ever-increasing speed and heavy usage to which the railway vehicles are subjected, imposes successively harsher condition in its components in the form of impact, rolling contact fatigue and thermal variations [17]. This elements are always under deterioration and their failure will lead to various problems. Their maintenance conflicts with time, economic growth and the need to compete with other markets [18].
Not only is intended to report the effects or symptomatic results, but it is also searched a response that is able to predict component failure and proactively schedule checks and repairs to the vehicle before they suffer extra damage, and stress track and structures [18]. There is a demand for an intelligent condition monitoring system that can respond to the cost efficiency, reliability and safety requests.

There are some assumptions that are erroneously taken for sure that were prove to be incorrect. Defects in wheel, bearing and suspension that are below industry accepted alarm threshold can be harmless, specially from a rolling stock perspective.
Historically, threshold were set based on levels determined to be harmful to track and structure. Systems that rely on this thresholds, "flagging" the vehicle as soon as this vehicles exceeds the determined levels, do not allow a reliable trending, have a low-resolution, have a high rate of false alarms and are inadequate and inappropriate for today's industry [19]. Trains must be often stopped in the network, pulled and inspected for defects, leading to expensive delay and increased repair costs. But above all, leaving components in service until they reach high alarm levels causes further cumulative and expensive damage to other components. On the other hand, premature replacement and repair of failing components is not desirable. There must be a critical examination if whether the fault types and fault levels can be left in service to maximise usable lifespan, without inducing collateral damage, or if they cannot, as the cost of waiting to repair the identified damage is significant [18]. For example, for a small defect that appear in

a wheel and grows over time, the adjacent wheelset will star to show signs of growing defects if the initial defective wheel is left in service. For longer service times, the wheels on the adjacent axles will start to deteriorate and in time the worsening of wheel defects will cause tracking error and subsequent wheel profile damage. Early maintenance of a modestly damaged wheel can prevent deeper and more costly repair. One should bear in mind that consequent caused damage is not the only factor that can represent costs: if equipment reach alarm levels on line, there is a considerable traffic delays related expense and high cost of preparing safe vehicles for safe transport to maintenance depots (for instane, a wheel or bearing that causes an alarm that requires cut-out in network costs approximately six to eight times as much as a repair performed in depot as part of predictive maintenance[18]); wheel impacts, aberrant wheel profiles, bearing damage and poor vehicle tracking geometry leads to increased fuel consumption.

There are various reasons to the implementation of a condition monitoring system up to the existing demands[4][17][20][21]:

- The early damage detection permits the reduction of overall maintenance costs, as a soon to be acute damage can be repaired in a early phase were its maintenance is less expensive, easier and less time consuming;
- It permits an optimization in maintenance schedules and reduces the need for a general regular maintenance;
- It permits a better management of spare parts logistic and optimization of associated costs;
- Possible losses are reduced, as the equipment isn't replaced for a new one in a healthy state;
- An up-to-date optimized functioning is possible at any given time, saving in operation costs as malfunctioning machine have a greater resources consuming rate;
- It improves the ride comfort, since there is a better monitoring of vibration, noise and temperature levels and a efficient diagnosis of its origin.
- Asset availability is improved, as service intervals are safely increased.
- It reduces the possibility of delays, as a "dead" train has implications for costumer perception, operator and manufacturer reputation, along with cost from penalties, recovery, train availability and repairs.
- It reduces the probability or derailments, reducing the risk to life and with the same repercussions of the previous topic but at a greater scale.

3.5 Conclusions

Nowadays, there is a great demand for railway organizations to deliver more while providing an affordable and high quality services. Recent legislation oblige the introduction of sustainable measures in their everyday services. All this challenges the railway industry to invest in the right decisions, optimising their activities.

Asset management plays a huge role in the achievement of successful results. One of the most important sections of asset management is asset polices. Although it covers quite diverse areas, we will address ourselves to the problem of inspections, maintenance and renewal intervention.

One of the most important and complex elements of a railway vehicle is the bogie. It is responsible for the correct movement of the train along the track and it strongly influences the ride comfort and stability. It is composed of diverse components, as the bogie frame, axle box suspension, axle bearings, wheels, axle, transmission, brake equipment, etc. The axle bearings

proves to be a critical component. It is not only the component with the highest failure rate (which is transversal to almost all kind of machinery) as its failure during the vehicles functioning may produce derailments, causalities, and high expenses. Since it is a component that mustn't run until failure, it is important to monitor its degradation state and to know the appropriated time to replace it.

For the correct monitoring of the bearing, the composition and properties of the existing bearing should be known. It is also important to acknowledge the possible failure modes, and also the axle ones, and how they affect the bearing. This information is crucial to the establishment of the main framework for the application of condition monitoring techniques.

Several techniques can be applied to monitoring of a railway axle bearing. In the following chapter, general considerations about condition monitoring will be made, followed by the description of some of the existing techniques for signal processing, diagnostic and prognostic of bearings.

Chapter 4

Condition Monitoring

4.1 Maintenance techniques

With the actual development of modern technology, the demand for the reliability of any type of machine is ever increasing, being required a precise and updated status of the machine condition at any given time. The condition of a machine and its performance depends mainly on the status of its components. Regardless of how well designed a component is, or the products and techniques used in it's manufacturing, every single one will deteriorate during its life span and eventually fail, if used long enough. There are so many factors that contribute to its deterioration, related to environment and mechanical properties and stress or strain pattern, that their exact evolution in never predictable: a relatively small change can have a huge impact during a long running time. Overall, randomness is always present and makes the exact component's condition impossible to know beforehand.

This means that for each machine, or a set of machines, a certain type of maintenance technique must be chosen. Maintenance is defined, according to EN 13306:2007, as the *combination of all technical, administrative and managerial actions during the life cycle of an item intended to retain it in, or restore it to, a state in which it can perform the required function.* Maintenance techniques can be classified into three different types: corrective maintenance, preventive maintenance and predictive maintenance (or condition based maintenance) [22].

One of the first maintenance technique being implemented was the breakdown maintenance (belonging to the corrective type). The machines were run until a breakdown occurred and only then the maintenance was made.
Following this technique, another was implemented, the time-based maintenance (that belongs to the preventive type). A time interval is defined, periodic maintenance is made at a this constant rate, regardless of the health status of the physical asset. With the successive developments in technology, machine elements became more complex and expensive, and the demand on machine's quality and reliability ordered smaller time intervals, proving the time-based preventive maintenance as an impracticable maintenance technique due to its high costs. The search for a new maintenance technique reappeared.

The ideal case scenario would be the knowledge of the exact condition of each component at each given time, and only act when they proved to be faulty. An approximation to this objective turned out to be possible with developments in technology, namely the creation of more precise sensors and better data acquisition systems, offering a wide range of possible solutions at lower costs.
The first concept of condition based maintenance was first introduced by the Rio Grande Railway Company, in the late 1940s, under the denominaton of predictive maintenance[22]. Since then, new maintenance techniques under the condition based maintenance category flourished and turned out to be the more reliable and cost effective. Today, they prove to be an indispensable requirement for the utilization of most of the existing machines.

4.1.1 Condition based maintenance versus Time based maintenance

As we will see, time based maintenance can be a better choice than condition based maintenance for certain cases.

Time based maintenance is easier to implement, once that the time that a unit has been on service it the only thing that must be recorded. However, substantial remaining useful life may be wasted if the asset is discarded still in reasonable condition, or breakdown might occur if it deteriorates faster than expected.

The benefits and success of condition based maintenance strongly depends on the behaviour of the deterioration process and the severity of failures. Some practical factors that affect the accuracy and precision of the collected data will severely affect the results of this maintenance technique.

Condition based maintenance should only be applied if the relative benefit (cost-savings) outweighs the efforts and costs of its application during the entire life cycle. To switch from time-based to condition-based maintenance is necessary to acquire and implement equipment and software that will store, analyse and initiate maintenance actions.

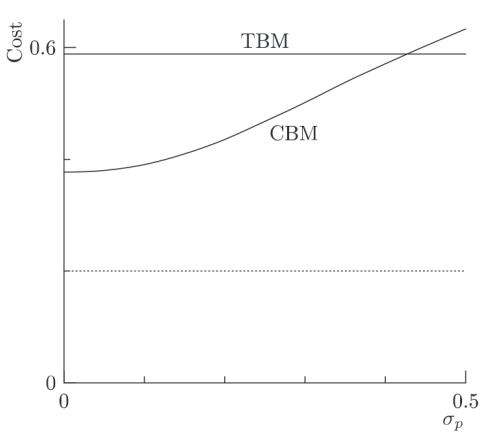
So, it may be important to understand if it justifies to switch from time based maintenance to condition based maintenance. Considering that both methods are working in optimal conditions, the performance of condition-based maintenance turns out to be much better than the time-based one, but in real case scenarios, it is no longer so. In some cases, a combination of this two techniques assures the best results.
De Jonge et al [23] studied how the influence o practical factors (planning time, imperfect condition information and uncertain failure level) can affect the benefit of condition based maintenance over time based maintenance. A succinct resume of the overall analysis will be presented.

There are some pre-established studies, [24] and [25], that support that the relative benefit of condition based strategy turns out to increase in the relative replacement cost and that cost savings are more significant for short times between inspections. So, there is a explicit relation between this technique benefits and the replacement cost and time between inspections.

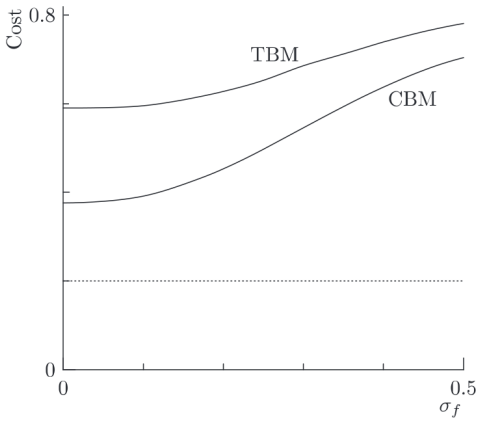
Other factors, as the behaviour of deterioration process and the cost of preventive maintenance, affect the overall benefit of one technique over the other. It turns out that the behaviour of the deterioration process is much more important than the cost of performing preventive maintenance. The cost differences between condition-based and time-based maintenance are substantial for small levels of variation in the deterioration process, but as this level gets bigger, the cost difference diminishes. Only for extremely small or extremely large preventive maintenance costs, the benefit of condition-based maintenance is limited; but for a wide range of preventive maintenance cost there are substantial cost saving.

The efficiency of both methods depend on various factors as, for example, the deviation of the deterioration process (the failure may occur sooner than expected/predicted). For condition based monitoring, the accuracy of the prediction is severely influenced by the quality of measured data. Vibration and oil analysis techniques are especially prone to inaccuracies. To study the influence of practical factor on the benefits of condition based monitoring over time based maintenance, we will consider three parameters at different "weights", input them in a mathematical model and draw results. The three parameters considered are the following: planning time, imperfect condition information and uncertain failure level.

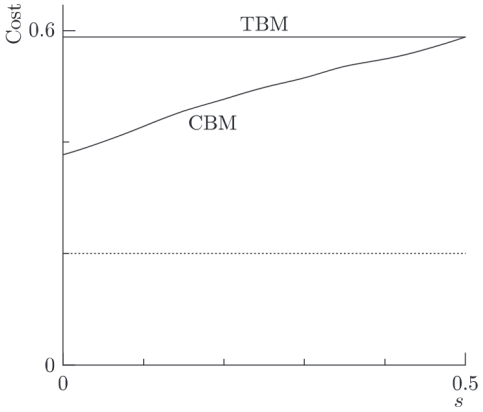
Planning time represent the "delay" between the time when it is realized that maintenance must be done (when condition-based monitoring detects a fault) and the time when maintenance is actually done (which is not always immediate, as the maintenance may depend on the availability of repair man or spare parts). It is considered that, for time based maintenance, it is always possible to perform the maintenance on the pre-defined data. *Imperfect condition information* represents the uncertainty in triggering the maintenance based on the acquire data. As this factor gets bigger, worse will be the results from the monitoring. As the measurement errors only influence the observed information, this factor doesn't affect time-based maintenance.



(a) Cost rate as a function of planning time s.



(b) Cost rate for a varying level of uncertainty σ_p in the obtained deterioration information.



(c) Cost rate for a varying level of uncertainty σ_f in the failure level.

Figure 4.1: Cost rate under optimal TBM strategy and CBM strategy.

Uncertainty failure level represents the deviation in the deterioration process. As it gets larger, the variance of the time until failure will also get bigger. It affect both condition-based and time-based maintenance.

Required planning time and imperfect condition information only affect condition-based maintenance whereas uncertainty in the failure level affect both of the techniques (although it has a stronger effect on condition-based than on time-based monitoring).
The cost benefit of condition-base over time-based maintenance decreases linearly in the planning time, until it equal zero when this value reach its limit (when planning time is equal the maintenance age of the optimal time-based maintenance strategy).
With imperfect condition monitoring, although for small values the cost benefit is substantial, for larger values there is a certain time where condition-based maintenance performs worse than time-based maintenance, turning out that the last one is more costly effective.
With uncertainty in the failure level, the relative cost benefit decreases, but it continues to be always positive.

All the factor can significantly affect the benefit of condition-based monitoring over time-based monitoring, although in the overall condition-based maintenance present a better performance than time-based maintenance.

4.2 Condition based maintenance

Unlike breakdown maintenance and preventive maintenance, the condition based maintenance focuses not only on fault detection and diagnostic of components, but also on degradation monitoring and failure prediction[22], allowing the minimal downtime and extended damages.

There can be found various definitions for condition based maintenance. According to British Standard BS/ISO 13379-1:2012, it is defined as *the maintenance policy carried out in response to a significant deterioration in a machine as indicated by a change in a monitored parameter of the macxhine condition.* According to an article published by Kothamasu et al. [26], it is defined as *a decision making strategy where the decision perform maintenance is reached by observing the condition of the system and/or its components.* Although defining clearly the goal of this type of monitoring technique, they fall shortly on including technical aspects, contrary to the definition given by Butcher [27], *set of maintenance actions based on real-time or near real-time assessment of equipment condition, which is obtained from embedded sensors and/or external tests and measurements taken by portable equipment.*

There are several advantages that justifies the use of condition based maintenance techniques [22]:

- It gives prior warning of impending failure, avoiding unexpected catastrophic failure;
- It can detect faults in early stages, enabling a proper planned maintenance. This is also useful for products safety management, as it can increase safety by detecting problems in advance before they occur;
- It enables a system to continue to operate as long as it performing is within predefined performance limits;
- It increases precision in failure prediction;
- It allows a better planned and less costly maintenance as it reduces unnecessary inspections and decreases time-based maintenance intervals with confidence;
- It aids in diagnostic procedures;
- It is attractive for maintenance service providers as a tool for monitoring the products quality during warranty period;
- It optimizes production processes and improves productivity;
- It provides a high-quality assurance, improving costumer satisfaction and avoids risk of cost due to the dissatisfaction of product quality.

Despite all of the benefits, condition monitoring as its drawbacks:

- It represents a high investment cost;
- It requires the installation and use of monitoring equipment;
- It requires the development of models and decision making strategy, which is especially long for new assets;
- It requires trained staff.

Condition monitoring can be a valuable tool that allows a considerable amount of savings, but its aplication should be well pondered. It will depend on the product type, its lifecycle and its importance for the overall system. As an illustration, in 2003 the annually savings due to the use of condition monitoring technology in the United States was estimated in 35 billion of

dollars [28] yet, in 2011, only 30% of all the industrial equipment benefited from it [29].
For large scale plants and high valued product, condition based maintenance is always advised as a product failure will cause great losses. For mass-consumption products it can not prove so effective in terms of maintenance cost.

The condition monitoring of bearings can be made based on the collection of different types of data. The most common types are: vibration, oil/debris analysis, temperature or acoustic emission [30].
Vibration is the most reliable and standardized method, is able to react immediately to change and detect the defective component. However its implementation is expensive and intrusive.
Oil analysis is able to detect the type and location of fault and easily characterizes the bearing condition. The required equipment is expensive and can only be applied to bearings with an oil supply system, as is difficult to apply on grease lubricated bearing and can't be applied to sealed bearings.
Temperature methods are standardized and are only able to detect anomalous working areas. It is difficult to detect impending failure since substantial rise in temperature only occurs in last stages of life, turning out to be ineffective in bearing prognosis and diagnosis.
Acoustic emission can detect impending faults, and if applied with care, may have a large signal-to-noise ratio. It is an expensive technique that requires a high sampling frequency.

Furthermore, target products are treated as an isolated unit system and not as part of an integrated system. For complex systems, it can be conducted at equipment level, one piece of equipment at a time.

4.2.1 Condition based maintenance procedure

Usually [31], condition based monitoring is done in the following order:

1. **Gathering of product status data and monitoring (data acquisition):** Definition of the type of data to be monitored (oil analysis, vibration, temperature, sound, etc) and which data should be gathered; Implementation of sensors (wire or wireless techniques); Connection of sensor to a proper database and monitoring system.

2. **Treatment of the given information (data process):** Signal processing techniques, for signal denoising, improvement of signal features, etc.

3. **Diagnosis (data analysis):** fault detection; fault isolation and determination of its location; fault identification and determination of fault mode. Diagnosis usually demands data pre/post processing, data interpreation and data fusion.

4. **Prognosis (data analysis):** estimation of time to failure; risk of failure mode and its probability; determination of possible future failure modes. It is usually predicted the remaining useful life and system health index are estimated.

5. **Maintenance action (maintenance decision support):** Execution of proper actions such as repair, left to use as it is and disposal

It is usually divided in four key steps [32], [31]:

Data acquisition

It is the first step of condition based maintenance, where the relevant information is collected and stored. The data acquired can be classified as event data or condition monitoring data. The first type of data informs about the occurrences happened and can enclosed the measures taken. The condition monitoring data comprises the measurements about the health condition

and the state of the physical asset. The data collected is extremely important (even event data, that is usually ignored). An updated and organized list of occurrences and measures taken can be a great help in future planning.
The information is collected through sensor and can contain values about the vibration, oil analysis, temperature, humidity, etc.

Data process

Data process allows the previously acquired information to be handled and analysed, permitting a better interpretation of the data. The data collected can be divided in three categories: *value type* (the data was collected at a specific time epoch and is described by a single value), *waveform type* (the data collected is a time series) and *Multidimensional type* (the data collected is multidimensional);
In this step, the data firstly undergoes a cleaning process (especially the event data) to eliminated possible errors, that can be originated by human input or sensor faults. Then, the data follows a second phase, data processing, also known as signal processing for waveform and multidimensional data. There are numerous different signal processing techniques that can be chosen. In the following section, signal processing for waveform data type will be listed (the type of data of interest for this document).

Data analysis

Techniques for data analysis are divided in two main categories, diagnostic and prognostic.
Diagnosis permits the identification of fault through the detection and isolation of a certain type of fault. When an abnormal response on the machine is detected, the fault is isolated from the signal and the nature of the fault and the faulty component are identified.
Prognosis allows the detection of the moment when the fault will happen and estimates how soon and how likely the fault will occur. Apart from the prediction of the residual life of the component, it can also estimate the evolution of wear.
Prognosis acts prior to the fail occurrence and has the main vantage of permitting zero-downtime performance. When prognostic methods are used, diagnostic ones should not be necessary, but both can be used at the same time, which is valuable for cases when the first method fails.

The correct diagnosis and prognosis of faults is closely dependent of an effective treatment of the vibration signal generated. The following requirements may be covered by the signal processing methods, but can also be integrated in data analysis:

- Correct extraction of the faulty signal:
  - Removal of noise;
  - Enrichment of the impulsive features;
  - High resolution in time frequency domain;
  - Ability of dealing with non-stationary and nonlinear character of the signal;
- Capacity to deal with real bearing functioning characteristics:
  - Robustness of the method under variable speed and operation conditions;
  - Removal of interference terms due to other vibration sources;
- Ability to give proper results:
  - Identification of the fault characteristic frequencies;
  - Extraction of fault features;
  - Estimation of defect size and time to impact;

- Signal processing requirements:
  - Minimal time consuming;
  - Simple in execution;
  - Fast to give diagnosis/prognosis results;

Although there is a wide number of existing methods for the monitoring of bearings, there are some areas that continue little explored: most of the fault analysis methods can focus only on a single point defect, don't account for the changing defect condition during the degradation phases and can't estimate the fault size.

Maintenance decision support

Maintenance decision support is the last step and the one that will permit to recommend efficient maintenance policies. Some decision issues should be established: determination of inspection frequency, establishment of warning limits, development of a decision method to select the best cost-effective maintenance operation (to understand which maintenance option is better under a given situation in terms of maintenance cost).

For example, sensor measurements can be taken at regular intervals or in a continuous, real time way, which seems the best way to analyse product status. If the monitoring is done continuously, it will lead to heavy load of data gathering which leads to high cost. It turns out to be non-cost-effective. It is best to base the data gathering on a certain time period (obviously its effectiveness will depend if the time interval is the most suitable and if the time period is adequate given the scale of gathered data).

For each option on what, when and how to do the maintenance, a construction of maintenance cost models must be constructed. Comparing different cost models is one of the best ways to select the best cost effective maintenance schedule. Action plans can also be design in consideration of the general framework that integrates data acquisition, data process, data analysis and maintenance decision support.

Table 4.1: The four key steps of the condition based maintenance procedure.

Data Acquisition	Data Process	Data Analysis	Decision Support
Type of data:	Techniques:	Diagnostic techniques:	- Maintenance models
- Event data	- Time domain	- Data-driven	- Action plans
- CM data	- Frequency domain	- Model-based	- Maintenance Schedule
	- Time-frequency domain		
Type of data:		Prognostic techniques:	
- Value		- Experience-based	
- Waveform		- Model-based	
- Multidimensional		- Data-driven	

4.3 Data Process

4.3.1 Signal Processing Techniques

There is a wide range of signal processing techniques that can be chosen for a certain application, and even some new ones can be designed from scratch. Some of the existing techniques can be merged for a best response about the system status.

Signal processing techniques for bearing monitoring will be synthesised in the present section. They will be divided in three groups, time domain methods, frequency domain methods and time-frequency methods [33], has shown in table 4.2.

Table 4.2: Resume of the signal processing techniques covered in this chapter.

Domain	Technique	Notes
Time	Statistical moments	Calculates statistical parameters that can be used to define thresholds or compare signals between healthy/non-healthy bearings.
	Autoregressive methods	The signal is modelled with an autoregressive models that can be used to forecast the values of the signal , so that deviation of the real signal can be detected. In enables the extraction of new signal features.
Frequency	Fast Fourier transform	It enables the transformation of the signal from the time domain to the frequency domain. It allows the output of a frequency spectrum, easing the detection of the more energetic frequencies. It is usually used as the first step of more complex techniques.
	Envelope Analysis	The effectiveness of this method depends on the proper selection of frequency bands. The signal is demodulated and, building a envelope spectrum, it is possible to detect the repetition frequency of the impulses from the signal.
	Spectral kurtosis and kurtogram	The fourth moment of the signal (Kurtosis) is calculated for each frequency. It allows the detection of impulsive signals and their frequency.
	Cepstrum	The spectrum detects the periodic components of the time domain periodicities, and cepstrum detects the periodicities in the spectrum. It is used to remove the periodic components of the desired frequency of the frequency spectrum.
Time-frequency	Wavelet Analysis	It permits the decomposition of the signal in a series of oscillatory functions, that can be shifted and dilated in time. It result can be considered as a set of impulse response of filters, allowing the detection of faults.
	Matching Pursuit	It decomposes the signal into a two dimensional space of time and frequency. It is possible to detect the impulses and their position in time and frequency.
	Empiric mode decomposition and Hilbert-Huang transform	The signal is decomposed into components called intrinsic mode functions, that contain different frequency bands. The Hilbert-Huang transform analysis these functions, identifies localized features and separates the contribution of different vibration sources.

It is not pretend an exhaustive description of all the existing methods, but a rough idea of some techniques employed in bearing diagnosis. A brief definition will be followed by the advantages/disadvantages and applications of each method. A more detailed review of the mentioned studies can be found at [33].

Time domain techniques

Time-domain techniques represent the earliest works on signal processing techniques and are, historically, the oldest method. Time domain analysis is based on time waveform itself and may involve the computation of descriptive statistical features such as crest factor, peak, mean, peak-to-peak interval, root mean square, standard deviation, kurtosis, skewness, etc. from the data. It is assumed that any fault can be monitored through this parameters. Crest factor and kurtosis is usually connected to incipient defects, whereas root mean square, associated with high energy levels, indicates the severity of defects [10]. The use of time-series can prove a superior technique, with a positive impact in the results. However, it is not suited for online monitoring, as it doesn't identify the defects responsible for the degradation and often generates false alarm for signals with a non-Gaussian probability distributions [34].
As a result of substantial speed variation in slow rotating bearings, the time domain is more appropriate for condition monitoring of slow rotation bearings. One of the most difficult challenges in these techniques is the extraction of the most useful information and link it to diagnosis or prognosis. Its accuracy strongly depends on the sensitivity and quality of the features extracted and selected to perform the monitoring.

There are various methods that can be used to perform a time-domain analysis:

Statistical parameters Simple signal analysis can be made based on calculated statistical parameters, such as kurtosis, root mean square (RMS) value, mean, peak, peak-to-peak interval, standard deviation, crest factor, etc.
As each fault generates its own acoustical signal, and symptoms can be found based on the amplitude and frequency of the signal. Unfortunately, the signal collect tends to contain noise and information about active vibration from other sources (mechanical components), that can mask the intended signal.
Some of the initially stated statistical parameters, especially kurtosis, crest factor and skew values can provide a good indication of the existence of faults, and the normalized skewness and normalized kurtosis prove to be useful in the detection of bearing fault at early stages. Statistical moments are insensitive to changes in load and speed condition. Despite this, a monitoring action based solely on these parameters proves to be too simple and non-effective, as signature analysis methods aren't very effective in minimizing the effect of noises and interferences.

Satistical moments techniques were the first to be employed in condition based monitoring. There are several studies based on this methods, prior to 2001, which were reviewed on [33]:
Root mean square values were successfully used to diagnose bearing failures. A threshold based on the value of RMS was defined, enabling the detection of high amplitude impacts of short duration;
A similar study was made, establishing the threshold based on the value of the kurtosis (and not on the value of RMS);
For the diagnose of bearing faults in early stage it was found that the value of normalized skewness and kurtosis are the best parameters to have into account. Results showed that these statistical moments are insensitive to changes in load and speed condition and can be used effectively;
A study compared the use of crest factor, kurtosis, skew and the use statistical parameters derived from beta distribution function in bearing diagnosis. It was found that the are no paramount advantages in utilizing the statistical parameters derived from beta distribution function on the detection and identification of bearing defects.

Autoregressive methods More advanced time-domain approaches can be used, as the application of time-series models to the waveform data. As waveform data is fitted into a parametric time series models, and features based on this model can be extracted. One of the most popular

model is the autoregressive model.
Autoregressive model can be used to model the vibration signal, forecasting its values based on past values. A change in machine condition may be detected by the comparison of the predicted signal and measured signal [35]. They are extremely effective when analysing short signals. Using this model is possible to generate a frequency spectra and extract information as, for example, its arithmetic mean, geometric mean, matched filter root mean square, root mean square of spectral differences, sum of square of differences that is more successful in classifying good and faulty bearing than the conventional Fast Fourier transform frequency spectra, in low-speed bearing applications. Applying this method to high frequency acoustic emission generated by the passage of the rolling elements in the damaged race, allows the examination of the health state of extremely slow roller bearings. An autoregressive moving average model was used with template-learning machine to enable the on-line identification of defect sensitive resonances.

Several studies, reviewed on [33], were focused on this technique.
A study reports that using autoregressive models to generate frequency spectra can be more successful in classifying good and faulty bearings than the conventional Fast Fourier transform frequency spectra, especially for low speed bearings and extremely-low speed bearings. It is effective when analysing short signals. It was also used in conjunction with Burg's algorithm and Akaike information criterion, obtaining a high-resolution capability allowing the detection of early defects, even with very approximate characteristic frequencies.
A technique called envelope autoregressive spectrum used the autoregressive model in conjunction with other techniques - adaptative network-predicated fuzzy interference fit, envelope analysis, band-pass filters - to diagnose the bearing condition, achieving significant gains in signal-to-noise ratio and a responsive and precise basis to trend bearing defects rigorously. A time-varying autoregressive method, contrary to other methods like short time Fourier transform, Wigner-Ville distribution and Choi-Williams distribution - can produce high resolution for the prediction of faulty bearing signals.

Frequency domain techniques

Frequency domain analysis involves the transference of the signal from the time domain to the frequency domain, allowing the extraction of certain frequency components of interest, after their isolation and identification. Examining the spectrum analysis, it is possible to look at the whole spectrum, or spectrum ranges, select the frequencies of interest, extract features of the signal and perform the derivation of the characteristics of bearing faults. The effectiveness of this method strongly depends on the operating conditions of the bearings, if they are stationary or non-stationary. A recorded bearing signal result on a mixture of different sources, corresponding to the external environment as well as nonstationary signals induced by internal components of ball bearing. In this case, it should be used a time-frequency method. [34].
One of the most widely used tool for the spectrum analysis is the fast Fourier transform.

Fast Fourier Transform One of the most widely used tool for the spectrum analysis is the fast Fourier transform, converting the time-domain content of the signal into a frequency spectrum. Although relatively easy and fast to apply, the defect frequencies are not easily distinguishable in the induced spectrum. In cases where the signal to noise ratio is low this technique is not successful, but a moving window technique can be used to improve this ratio to more acceptable levels. So, it is not a good choice to make a correct diagnostic decision if used alone, since time domain features of the signal are ignored.

When using a Hilbert-Huang transform prior to the fast Fourier transform analysis, fault frequencies can be accurately captured.
Other Fourier transforms can be used, as, for example, the pseudo-Wigner distribution. It

is applied in the analysis of bearings when exists a stationary component at the fundamental impulse repetition frequency of the signal.

Some incipient spectral implement such as bispectrum and power spectral density can be used to compensate for the deficiencies of the fast Fourier transform technique.

Envelope analysis methods The envelope analysis is based on the demodulation of high frequency resonance associated with bearing element impacts. It is dependent of proper selection of frequency bands for demodulation and the detection of the pretended frequencies. It typically contains clearer fault information than the raw signal. Analysing the envelope spectrum of a bearing signal, the repetition frequency of the impulse response series is easily spotted [8].
This technique can be extended by normalizing the frequency spectra, producing a very sensitive measure of the defect frequency.

In an experimental approach, an enveloped estimation algorithm was developed based on the resonance frequencies of bearings. It permitted the exclusion of noise, recovery of envelope signals matching their resonance mode vibrations, less distortion, low sampling speed and less computation loading.

Spectral Kurtosis and Kurtogram The spectral kurtosis can be defined as a statistical variable based on the fourth moment of the signal. It can be used to calculate the kurtosis of each frequency line in a time-frequency diagram, or be used to select the most impulsive parts of the signal [36]. It is useful to monitor parameter in situations where pulses contribute significantly to the signal level, although it can't handle excessive background noise and is vulnerable to false vibrations.

Used in combination with other techniques, it can show improved results:

- Combined with the autoregressive model and the minimum entropy deconvolution technique, the defect recognition of the spectral kurtosis and the envelope detection is improved, proving to be useful for the prognosis of bearing defects.

- Using this method based on a window superposition technique, amplitude information is better conserved.

The kurtogram is a diagram representation of various values of the spectral kurtosis, calculated by a series of filters with different parameters. Contrary to envelope analysis methods, this permits the utilization of band pass filters with non-fixed parameters of bandwidth and centre frequency, enabling this technique to estimate faults under varying speed and fault conditions.

Kurtogram based on short time Fourier transform or FIR filters should not be used because they don't turn out to be precise.
Alternatively, a kurtogram based on Daubechies-wavelet packet transform can be used, as it is more effective in de-noising de signal and extracting faulty features, and computationally simpler and faster than the original method.
Swapping the temporal kurtosis of the signals extracted from the wavelet packet nodes with the kurtosis of the corresponding power spectrum of their envelopes, improves its ability to restrain heavy noised and its visual examination potential.

Cepstrum Cepstrum is the anagram of spectrum, and can be viewed as the spectrum of the spectrum. It is represented as a function of an independent variable quefrequency, having the dimension of time.

Despite the many improvements applied over time to adapt this technique to rolling bearings fault investigation, its application to diagnose bearings is still ineffective in the presence of excessive background noise. It can be useful in the pronostic of bearing faults as, with the

minimum variance, it is even possible to predict fault periods regardless of system frequency response or choice of optimal resonance bands.

A study was conducted and concluded that using cepstrum analysis on a processed, normalized and integrated pre-whitened signal showed good results in the calculation of spall size and time to impact. Another study concluded that it is less complicated to apply cepstral based techniques then spectral kurtosis based methods for detecting bearing faults. Unlike spectral kurtosis methods, cepstral based techniques allows the demodulation of the entire band without the choice of an optimum band, whithout any loss of information in fault condition. A detailed review of both studies can be found in [33].

Time-frequency domain techniques

Contrary to frequency domain techniques, time-frequency techniques can handle non-stationary waveform signals. They used time-frequency distributions, representing energy in a two dimensional function of time and frequency, revealing more clearly fault patterns. Contrary to time and frequency domain methods, this techniques have the ability to deal with signal that result in dynamic behaviour of stationary/nonstationary vibration signals mixed with brackground noise.

The most simple used techniques are the Wigner-Ville distribution, Short-time Fourier transform and wavelet analysis. However, this techniques have their downsides: short-time Fourier transform have a constant time and frequency resolution, the Wigner-Ville distribution can lead to the appearance of cross terms that lead to misinterpretation of the signal and the wavelet analysis strongly depends on the quality of the analysed signal [34].

Wavelet analysis A wavelet transform is a time-scale representation of a signal. It allows the wavelet signal to be expressed in a series of oscillatory functions with different frequencies, and at different time dilatations, depending on a set of parameters previously chosen.
One of the main advantages of wavelet transforms is their capacity to produce a high frequency resolution at low frequencies and a high time resolution at high frequencies, and can reduce noise in raw signal.
In bearing analysis application, discrete wavelet transform is efficient in detecting single and multiple faults in ball bearing. Contrary to other methods as envelope techniques, it can remain sensitive to detect flatten flaws for longer periods. Using this aproach at Mel frequency scales is possible to reach a detection rate as high as 99% with only a minute window of data.
Used in combination with multi-resolution analysis, as it is very efficient in extracting information from chosen frequency bands, wavelet transform proves to be useful in diagnosing bearing fluting.

Wavelet functions can be used to produce filters. In most of the cases, the mother wavelets used in wavelet filters are chosen among an existing family of wavelets, with great care, as a wrong selection can lead to errors. But in other cases, the wavelet function can be constructed directly on the bearing signal, leading to successful results. There was a study where an antisymmetric real Laplace wavelet filter was constructed based on the signal an lead to a more sparse representation with an effective enrichment of the impulse attributes of the de-noised signal.
This technique suffers from the problem of selection of proper frequency bands for demodulation and the detection of the pretended frequencies.

Wavelet packet decomposition can attain higher descrimination compared to normal wavelet analysis, by analyzing the higher frequency domains of a signal. Frequency domains are easy separated by the wavelet packet, being later selected and classified acording to the characteristic of the analyzed signal. It is more appropriated than wavelet in signal analysis: it has much wider applications such as signal and image compression, de-noising and speech coding [37].

The article [33] reviewed various studies that focus on wavelet analysis: A study was made to understand the restrains of well-known envelope techniques used on bearing faults evolution process, comparing them to the wavelet transform method. This last method, contrary to other envelope techniques, remains sensitive to defect for a longer duration, even when the flaw is flatten by the contact between the bearing surfaces. Other study showed that wavelet transform can diagnose bearing fluting ageing when used in conjunction with multi-resolution analysis.
Discrete wavelet transform can be efficiently implement for the detection of single and multiple faults in ball bearings, whether they are located in the its inner or outer race. Another study used this method at Mel frequency scales. Result showed that the detection rate reached as high as 99% even with a minute window dimension training data.
Wavelet approaches can also be used to de-noise bearing signals. A study using a Mayer wavelet and the Weibull negative log-likelihood function were used to train a support vector machine classifier that was 100% efficient in fault discovery with bare least amount of input features. Wavelet filters were also used in conjunction with spectral subtraction to de-noise the signal and where effective in diagnosing rolling element bearings.

Matching pursuit Matching pursuit technique decomposes a signal into time frequency functions (called atoms) localized both in time and frequency. This technique proves to provide high signal to noise ratio, with a better performance than continuous wavelet transform and envelope detection techniques.
One variant of this technique, called basis pursuit, decomposes the signal based on different factors of the matching pursuits decomposition. It presented improved results, at a finer resolution, higher signal to noise ratio and, consequently, a more accurate diagnosis. It requires shorter lengths of data but takes a longer computation time than the previous technique.

A study was made employing matching pursuit with time frequency atoms to analyse and extract vibration signatures. It provided a high signal to nose ratio, with a better performance than the continuous wavelet transform and some envelope detection techniques (such as Hilbert transform) in the premature detection of faults. Another study used this technique to quantify the spall size fault and increment the precision of the defect sighting, using an impulse dictionary. The impulse dictionary should contain a quantitative connection among the step created in the signal by the entry and exit of the rolling element through the fault and the fault size. A detailed review of both studies can be found in [33].

Empirical mode decomposition and Hilbert-Huang transform Empirical mode decomposition is based on the local characteristic time scales of a signal and can decompose the signal into a set of complete and almost orthogonal components called intrinsic mode function, which indicate the natural oscillatory mode imbedded in the signal [38].
It is a technique that is suitable for non-linear and non-stationary processes, although it has been reported to introduce mode mixing effect and distortion of the faulty impulses.

It is usually used in conjunction with Hilbert-Huang transform. Hilbert-Huang transform examines the intrinsic mode function and identifies localized features, being able to separate the contribution of the different vibration sources. Its utilization is valuable for cases of shock and resonance that causes choking low frequencies and prevents early detection of faults.
Local mean decomposition is a new adaptative non-stationary signal method for feature extraction, that is based on the fundamentals of empirical mode decomposition. It shows a better performance in maintaining the information integrity of the signal and a lower number of iterations when compared to empirical mode decomposition and wavelet performance [7].

There were several studies developed analysing the advantages from conjugating this to other

known methods.
Used in conjunction of a morphological filter tuned to the right values, the entire impulsive signal can be extracted from background noise. For example, after applying an ensemble empirical mode decomposition to a spectral kurtosis optimized band pass filtered signal, it was possible to obtain a high number of intrinsic mode functions that lead to a successfully separation of the bearing signal from heavy nose.
Merging the empirical mode decomposition method with the approximate entropy method allows for a clear distinction on the entry and exit events, allowing the estimation the defect size; merging it with the bispectrum method it is possible suppress residue noise in the signal building phase and achieve high computation speed [33].

4.3.2 Denoising techniques examples

Contrary to what is desired, the acquired signal from an active bearing contains not only the impulses generated by the flaws but a combination of these impulses and heavy noise. The heavy noises mask the intended information, complicating the identification of bearing fault frequencies, especially when the signal to noise ratio is low. Some techniques were used to circumvent this problem:

- Using a conjunction of lest mean adaptive filter and Weiner filter it is possible to improve about 32 dB in the signal to noise ratio.
- Adaptative Self-Tuning filter provides a good incrementation of signal to noise ratio.
- Spectral subtraction based on the short time fourier transform enhances the sensitivity of temporal indicators, enabling the eraly detection of defects.
- Using the discrete Mayer wavelet and the the Weibull negative log-likelihood function is possible to denoise the signal allowing a great posterior precision in the discovery of faults.
- Wavelet de-noising method is performs very well on Gaussian noise and can almost achieve optimal noise reduction without altering the signal properties. Is involves three steps: signal decomposition, threshold detail coefficients and signal reconstruction. It relies on the fact that the energy of a signal will often concentrate in a few coefficients in the wavelet domain, being affected by signal with a lot of impulse components.

4.4 Data Analysis

Data analysis methods can be divided in two big groups, diagnostic and prognostic. But each groups contains several techniques that can be classified according to the specifications of the method used.

4.4.1 Failure Diagnostic Techniques

Although diagnostic is a more mature field than prognostic, the complex inter-disciplinary expertise required for the use of information from diverse domains (machinery dynamics, signal processing, statistical analysis, machine intelligence, etc) leads to a continual need of paradigms for diagnostic.

Diagnostic involves the mapping (or pattern recognition) of the data measured and features extracted from the failure space. The data is classified based on the extracted features. The effective evaluation of the degradation process proves to be difficult has each feature is only effective for depicting specific defects at specific stages. The quantitatively evaluation of the fault severity, especially at an early stage of the fault, is a difficult process.

Diagnosis methods can be divided into two main denominations: data-driven diagnostic methods and model-based diagnostic methods.

Table 4.3: Type of failure diagnostic techniques

Type	Methods
Data-driven	Artificial intelligence approaches Statistical methods Damage quantification indexes
Model-based	Quantitative models Qualitative models

Data-driven diagnostic methods

Data-driven diagnostic methods can be divided into three different approaches [31]

Artificial intelligence approaches This approach aims for an automatic identification of a certain type of fault, in which the system has the ability to continuously upgrade and expand its data base. The lack of proficient algorithms for acquiring training data forces the use of simpler systems, where measured data is used to train the system.

The most used method is the artificial neural network. Processing elements, made up of weights and nodes, are linked in a complex layered configuration. In the training process the weight are adjusted through the input and output observations, bringing the non-linear test function as close as possible to the real. If properly tuned, this complex layered configuration enables the model to estimate the state of health of a machine.

There are other well-known used artificial intelligence methods as fuzzy-neural networks, fuzzy logic systems, evolutionary algorithms and neural-fuzzy systems.

Statistical methods Statistical methods involves the use of statistical approaches, and the utilization of traditional statistics parameters, as crest factor, peak, mean, peak-to-peak interval, root mean square, standard deviation, kurtosis, skewness, among others. The principal difficulty in the development of this method is the proper selection of the most important information from the large dimension of initial features. It is also difficult to incorporate all the advantages from the extracted ones. The accuracy of the fault detection will depend on the quality and sensitivity of the original signal.

The Bayesian approach can also be used. The fault detection is estimated by a probability based on two antecedents, a prior probability estimation and a likelihood function based on a probability model of the observed data.

Damage quantification indexes Damage quantification indexes provides a quantifying degradation indicator for the assessment of machine performance. Efforts are being made so that this methods can handle various signal-related problems, as noise filtering, non-stationarity and uncertainties in the data, nonlinearities and multimodal distribution of the extracted features, noise-filtering, and different operating conditions.

Various methods can be used: self-organizing maps that lead to the calculation of a minimum quantification error; methods based on Gaussian order statistics; sum of N condition indicators; normalised energy; fuzzy support vector data description. These methods produce results that better reflect the different damage growth than methods based on RMS, kurtosis and crest factor.

Model-based diagnostic methods

Model based diagnostic methods locate the malfunctioning component of a system solely on the basis of its structure and behaviour. The approach can be made based on analytical approaches, using **quantitative models** or **qualitative models** to construct a knowledge-based approach.

In every method a precise mathematical and physics definite model is established, which is used in combination with the data collected to obtain residual data through residual generation techniques. Comparing the expected results with the data acquired, it is possible to diagnose the functioning machine. Usually, the systems are too complex and the establishment of a mathematical model is problematic or not practicable.

4.4.2 Failure Prognostic Techniques

Prognostic is one of the most useful tools in condition monitoring, predicting the failures before it occurs or before the monitored parameters reach a certain threshold. This proves to be extremely convenient, avoiding for the asset to undergoes the fault. It also prevents that when a certain damage level is reached and the maintenance is urgent, there are no spare parts or the maintenance equip is unavailable.

There are various definitions that can be found for prognostic in the literature, but according to the International Standard Organization (ISO) it can be defined as *the estimation of the Time To Failure (ETTF) and the risk of existence or later appearance of one or more failure modes.* Most of the literature uses the terminology Remaining Useful Life (RUL) insteade of estimation of Time Until Failure. So, apart from predicting the future health state of a given component, estimating the Remaining Useful Life is one of prognostics principal aims [37].

The first classification proposed for prognostic technique was a classification of three main approaches: model-based, experience-based and data-driven [30]. This classification method was created according to four criteria: complexity, cost, precision and applicability. It is the most used classification system and will be the one used in this report.

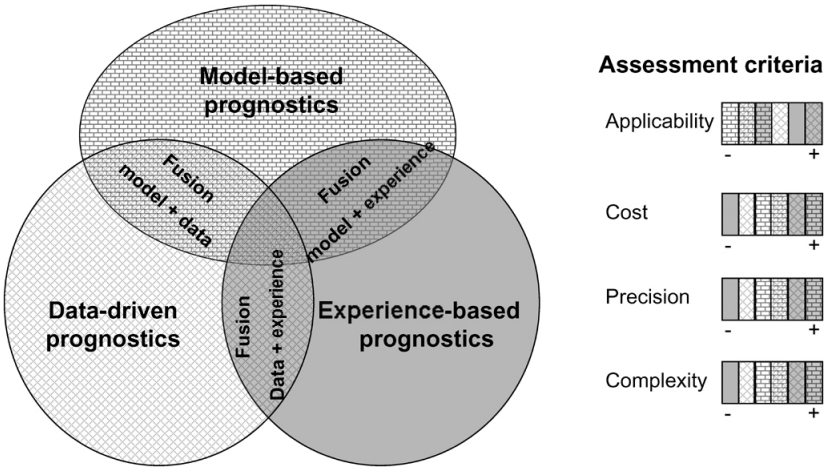


Figure 4.2: Classification system of fialure prognostic techniques [37].

Alternatively, it could be chosen the other classification system, that divides the techniques into model-based (physics of failure), data driven and hybrid approaches. As the experience-based approach used data collected from the experience feedback, it can be put within data-driven approaches. So, in this classification, the experience-based and data-driven techniques are merged into a single group, also called data-driven, and a new group was created, the hybrid approaches.

Table 4.4: Failure prognostic techniques covered in this chapter

Type	Method	
Experience-based	Expert Systems	
	Fuzzy Systems	
Model-based	Paris-law crack growth model	
	Formam-law crack growth model	
	Fatigue spall model	
	Contact analysis model	
	Stiffness-based damage rule model	
Data-driven	Statistical approaches	Trend projection model
		Particle Filtering
		Regression Analysis
		Dynamic Bayesian network
	Machine learning approaches	Hidden Markov model
		Artificial neural network (ANN)

Usually, the prognostic procedure is carried in the following way: the defect is detected in an early stage; the defective system is assessed and tracked continuously; a prediction with a confidence interval is generated, estimating the remaining useful life and possible failure modes. In bearing prognostic, signal modulation and noise are two major barriers, making the fault difficult to detect.

Experience-based prognostic methods

The time to failure and remaining useful life is predicted based on various parameters collected over a significant period of time (as failures times, maintenance and operating data) that allows the estimation of parameters of well known reliability laws. The experience feedback collected provides a computation straightforwardness that replaces the use of advance and complex models, and the results are extracted using simple reliability functions, as Exponential law, Weibull law, etc. Weibull law is the most reported one in literature, as it can represent several time phases of a component's life. The methods are based under average conditions an can lead to incorrect estimations during periods of changing operating conditions.

Two types of systems can be established for this type of methods: expert systems and fuzzy systems. Expert systems are simple, being easy to develop and to comprehend, although they need a substantial number of rules, precise and exact input data, are completely dependent on an actual and updated knowledge of the subject monitored. They end up not providing a precise estimation of the remaining useful life or supplying a confidence interval. Fuzzy systems, on the contrary, need less rules than the expert systems, can respond to inaccurate, noisy and incomplete input, and provide confidence intervals for some type of models.

These methods, although less expensive to implement, prove to be less precise when compared with model-based prognostic and data-driven methods, especially on new systems where lacks existence of experimental data and on systems where the operating condition varies frequently. It should not be used in system were the prognostic results are critical.

Model-based prognostic methods

These methods estimates the health state through an analytical model that represent the dynamic behaviour of the system and its degradation phenomenon. It can be constructed using differential or algebraic equation based on expertise know-how, theories and physic laws relevant to the monitored machine. One of the most challenging issue is defining the loading-damage relationship and to modelling it. These methods are usually based on simplified rules as linear damage rule, damage curve rule or double-linear damage rule, which have the drawback of considering the damage factor as constant, which is hard to estimate or measure.

This method consist of two different phases: in the first one, residuals that represent the dispersion of sensed measurements and the ones estimated from a healthy system are employed; in the second one, the failure degradation is mathematically modelled. Various modelling techniques can be used to create a model of a real system, usually about crack by fatigue, wearing and corrosion phenomena. For example:

- Paris law crack growth modelling: the performance relies on the correctness of the crack size estimated through the vibration data and the calculation is dependent of the bearing geometry, defect size, load and speed;
- Forman law crack growth modelling: relies on condition data and crack growth physics, but has to be further tested to prove its accuracy an applicability;
- Fatigue spall model (initiaton model and progression model): time to spall initiation or since spall initiation to failure is calculated, considering cumulative damage in conjunction with the operation condition, although there are several physics parameters that need to be previously determined;
- Contact analysis for bearing prognostic: finite element analysis is used to calculate stress based on the defect size and the bearing geometry, size and speed, having the downside of being computationally expensive;
- Stiffness-based damage rule model: establishes a relation between the component natural frequency and amplitude to the running time and failure time, needing previously that several material constants are determined.

The degradation phenomenon is represented by one or more variables and the dynamic is imposed by a set of parameters that depends on the environment within the physical system evolves. In some cases, the degradation variables can be considered part of the global behaviour model.

For extremely simple systems this method proves to be a good alternative, but as a system gets more and more complex, it is difficult to obtain an analytical model that expresses the non-linearity of the system and the generally stochastic degradation mechanisms. Apart from this, in industrial applications, the defect type usually varies from asset to asset, proving to be difficult to identify faults without interfering with the operations. Although it is more accurate in predicting results than experience-based methods, the applicability of model-based methods is limited.

Data-driven prognostic methods

Data-driven methods are based on the transformation of monitoring data into behavioural models, using artificial intelligence tools. Relevant information converted from on-line data captured by sensors, estimating the health state of the system. The degradation model is constructed upon different statistical models, tools or methods that don't take analytical models or physical parameters into account.
Data-driven methods can be sub-classified in statistical approaches and machine learning approaches. In the first, models are constructed fitting a probabilistic model to the available data.

In the second, the system tries to recognize complex patterns, using the empirical data for intelligent decision making. Both approaches are a two-step method: a first phase, during which the behaviour model is learned, and a second in which the learned model is used to estimate the current operating condition of the system and predict its future state. The need for samples from various equipment failure progression states turns out to be a downside as in some industrial systems the machines are not allowed to run until failure.

Methods can be constructed based on the following models:

Statistical approaches

- Trend projection models: are one of the simpler approaches, providing easy calculations, although they often lead to inaccurate forecast under periods of change in working condition as they are based on past degradation patterns;
- Particle filtering: can provide nonlinear projections but its performance decreases with the increase of the data dimension;
- Regression analysis: cannot provide a indication of time to failure of probability of failure. Used in conjunction with fuzzy logic it shows improved results. Regression analysis highlights the most recent condition information and fuzzy logic classifies it based on previous occurrences.
- Dynamic Bayesian networks: contrary to almost all of the previous methods, it provides a good estimation based solely on the condition data, rather than event data, but its accuracy depends of the previously correct determination of thresholds for several features

Machine learning approaches

- Hidden Markov model and Hidden semi-Markov model can be trained to recognise different bearing fault types and states but the prognostic projections rely on failure thresholds. It is used to model degradation in bearings and estimate the underlying remaining useful life.
- Artificial neural networks(ANN) is used for different applications. It can be used to estimated crack propagation of a bearing (track the time evolution of a crack size, and estimate the value of the remaining useful life). Time series estimation using ANNs do not require previous knowledge, are fast in handling multivariate analysis and can provide nonlinear projections, but have a short prediction horizon, assuming that the failure occurs once the condition indexes (which are assumed to represent the actual asset health) surpasses a presumed threshold. Exponential projection using ANN and Data interpolation using ANN have a longer prediction horizon but need an independent ANN for each monitored machine. Exponential projection has the additional advantage of predicting the actual failure time, but assume that all the degradation processes follow an exponential pattern.

The recent development of sensors and computer science has facilitated the development of artificial intelligence, and thus, of data-driven methods. These have two advantage over the previous methods: real monitoring data leads to more precise prognostic results; and getting reliable data in real applications is easier than constructing physical or analytical behaviour models. To construct a behavioural model of a bearing's degradation is very difficult, even impossible, and some already existing models are only valid for a specific bearing in specific conditions- Thus, data-driven methods can be considered a good alternative for bearing applications, although they give less precise results than model-based methods (but better than experience-based ones). They are less complex and more applicable than model-based methods, although they have a potentially long learning time.

There are some future challenges that can be developed in data driven methods for the calculation of the remaining useful life: models based on very few or no data situations, easing their application on newly commissioned systems without event data; data fusion of multi-dimensional input data from condition monitoring; develop a consistent way to model the influence of external environmental variables; development of models which can deal with multiple failure modes [39].

Chapter 5

Experimental test

The UTE 2240 unit is an electric multiple unit used by Comboios de Portugal (CP). In an attempt to modernise the older vehicles, the structure of the existing models of 2100 and 2200 series were used, and the interior and all electrical system were rebuilt. The original mechanical system was built by Sorefame industry and the posterior rehabilitation performed by Alstom.

Figure 5.1: A photo of the UTE 2240 (yellow model).

It is composed of two non-powered cars at both ends, and a powered car in the middle. It has a two-axle, non-articulated bogie, meaning that for each car we have two pairs of wheelsets. The total unit as a total of 24 axle-bearings.

The axle-bearing in use is the SKF's Tappered Roller Bearing (TBU) *1639590 A (130 × 220 × 150)*.

It is constructed from the SKF bearing *BT2B 641162 C*, a double-row tapered roller bearing, which has a single outer race and two inner races, separated by a spacer ring. The bearing elements have two rows of rolling elements with the shape of an truncated cone, two individual cages, and is sealed at the extremities. On the inner side, the ring has a seal outer ring and a seal inner ring equipped with a Z labyrinth, and a distance ring to prevent lateral wear against the axle. On the outer side, it has a provisional distance ring and a distance ring. There is a threaded plug on the outer race that permits the adjustment of the grease levels in the bearing's interior.

The actual maintenance technique in use for these roller bearings is a time-based one. The bearings are replaced at a constant time interval of what is estimated to be the useful life of the bearing reduced by a large safety coefficient, imposed for critical application. These bearings are specially developed for railway application and have a large acquisition cost. Replacing the bearing in a healthy state represent a cost loss, as they are discarded while still in an acceptable condition.

As we have previously seen in section 4.1.1, condition-based maintenance prove to be more cost-effective than time-based maintenance. The application of a condition-based maintenance

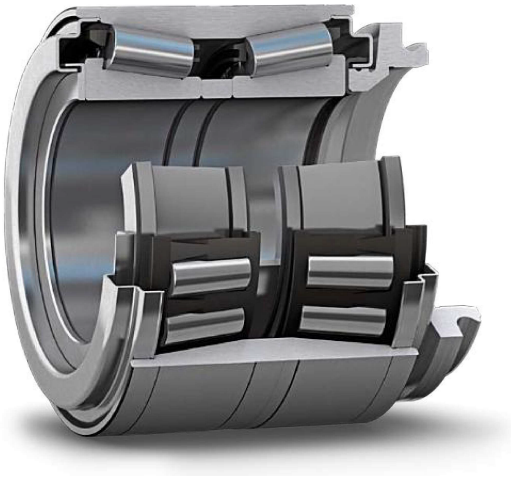


Figure 5.2: Model of tapered rolling bearing unit (TBU) similar to the rolling bearing TBU 1639590 A (130 × 220 × 150) [6].

should then be considered.

The condition-based maintenance can be made based on different physical parameters of the asset. For bearings, there are usually four types of parameters that are monitored: vibration, lubricant analysis, temperature and acoustic emission. The analysis of temperature is one of the less costly techniques, but it is not suitable for the detection of impending faults since substantial rises in temperature only occurs in last stages of life. Since this is a critical application temperature monitoring should only be used as a secondary technique to complement other that was implemented.
Oil analysis techniques, although expensive, are able to detect the type and location of the failure. However, they are almost limited to bearings with an oil supply system. In the case of UTE 2240's units, the bearing is grease-lubricated, and the analysis involves the withdraw of the train from the tracks and its partial dismount, which results in elevated costs. Plus, the bearing is sealed and lubricated for life, which make the application of this type of techniques impossible.
The techniques for acoustic emission and vibration are very symilar, although they differ in the measuring process. Most of time, for acoustic emission monitoring, the microphones are located on the track and the vehicle must pass through it at a certain speed for collection of data. Vibration, although more intrusive, is more flexible as the accelerometer are incorporated in the train. In overall, acoustic emission has a large signal-to-noise ratio and only impending fault can be detected, whereas vibration is a reliable and standardized method, as it reacts immediately to change and can point out the defective components [30].

Developing a condition-based maintenance to monitor these roller bearing is a laborious and complex process. Some previous studies must be done so that the system properties and general behaviour may be known. As it was already mentioned before, is very difficult, if not impossible, to model a bearing and often there are considerable deviation between the expected and the real, measured behaviour.

One of the best ways to collect event data is to utilise a test bench to collect general information that will be useful for future development of a monitoring system. Its utilization will permits the extraction of the kinematic behaviour of the bearing, as well as the categorization of the bearing's response (with and without the most usual defects), for different loads and speeds.

5.1 Test bench

Integrated in the *SIEF Project* (Integrated system for reliability of railway equipment Project), a test bench was constructed between the year 2011 and 2012, developed specially for the UTE 2240's axle bearing. It can be encountered in the facilities of the Faculty of Engineering of University of Porto (FEUP), in the laboratories of building M.

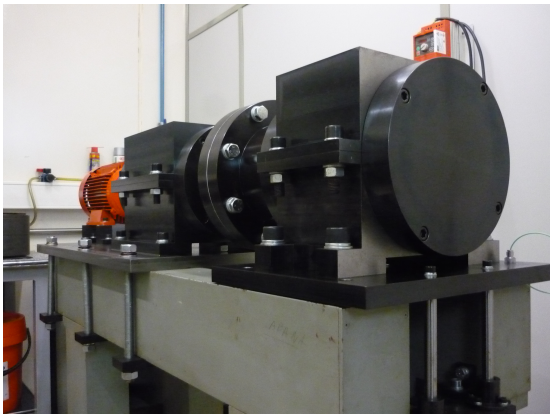


Figure 5.3: Photograph showing the existing test bench.

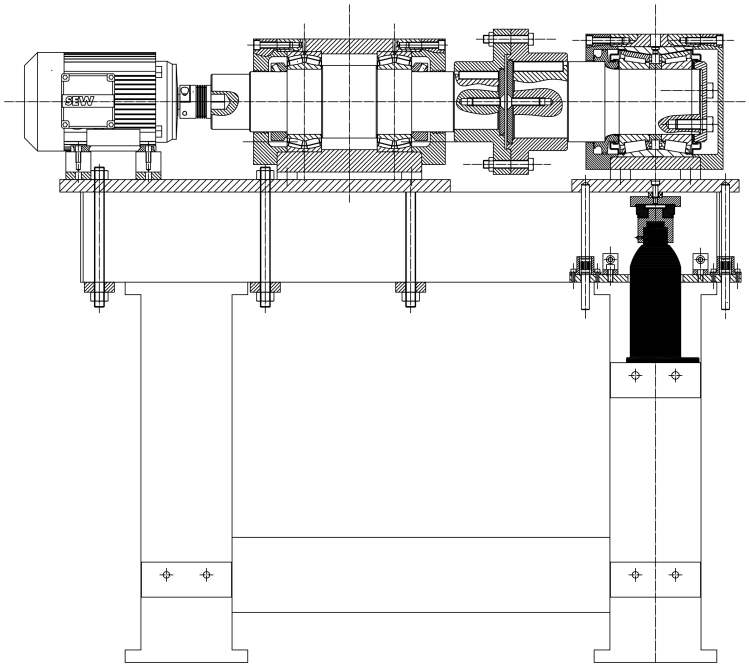


Figure 5.4: Schematics of the test bench.

It enables the collection of the bearings vibrational data through a recreation of its functioning.
The test bench is powered by an AC electric motor, *SEW DTL100L4 (B3)* that is connected to the extremity of an axle through a precision coupling, *R+W BK7*. The axle is supported by two spherical roller bearing, *SKF 24026 CC/W33*, inserted in a box located close to the motor. Both the motor and the box containing the supporting bearings are attached to a support through the use of threaded rods and nuts.
The axle bearing that will be monitored is located on the other end of the axle, inside a box,

in which is applied a load through the use of a hydraulic jack, *LUNA 20796-0501*. This box can move freely and is only guided vertically by four pairs of a solid shafts *INA FAG W16* and a linear ball bearing *INA FAG KH16-PP*. The applied load can be measured through a force sensor *Kistler Load Washer 9103A*, located between the axle-bearing box and a thrust spherical plain bearing, *SKF GX30F*, connected to the hydraulic jack. The vibrational data can be collected through the use of accelerometers, that can be fixed to the support, axle-bearing box or directly to the outer race of the bearing, accessible through two slots cut into the top and side of the mentioned box (the radial vibration can be acquired in two perpendicular directions), as shown in figure .

(a) Through the top of the box.

(b) Through the lateral side of the box

Figure 5.5: Access to the outer race of the bearing.

5.2 Acquisition of data

5.2.1 Force acquisition equipment

For measuring the load applied by the hydraulic jack to the axle-bearing box, it is necessary to receive the outputs from the force sensor. The signal sent from the sensor must be amplified. The charge amplifier in use is the *Kistler ICAM 5073A*, developed for various Kistler force sensors, one of which is in use on the test bench, the *Kistler Load Washer 9103A*.

The charge amplifier has 3 different connectors: a BNC connector, D-Sub 15-pin male connector and a D-Sub 09 female connector.

Figure 5.6: Kistler ICAM 5073A.

The sensor cable should be connected to the BNC jack.
The power supply of the amplifier is done through the D-Sub 15-pin connector (18-30 VDC), so

this connector should necessarily be used. Apart from the power supply, this connector enables the input of controls and the output of the analogue measured signal. Each pin of the connector can be utilized for different commands. A detailed description of the pins allocation, along with the corresponding colour code of the available cable in the laboratory, can be found in table 5.1. The D-Sub 9-pin connector can connected to a PC for parameter settings via the RS-232C interface, through the *ManuWare* distributed by *Kistler*. It is also possible to start or reset measures, export the latest measured values and peak values, and output a measured value continuously.

Table 5.1: Pins allocation and corresponding cable colour for the D-Sub 15-pin interface.

Pin	Allocation	Cable colour
1	-	orange
2	-	red
3	output channel 1	blue
4	-	white
5	-	green
6	peak of channel 1	black
7	common Control	red (white stripes)
8	measure	blue (white stripes)
9	power supply (ground)	green (white stripes)
10	output channel 1 (ground)	blue (black stripes)
11	power supply (+18-30 VDC)	black (white stripes)
12	-	orange (black stripes)
13	-	green (black stripes)
14	-	red (black stripes)
15	range of channel 1	white (black stripes)

As the load applied to the bearing will be constant during the experimental test, and in an attempt to simplify the process, we will utilized the D-Sub 15-pin connector only for power supply. The outputted signal will be visualised directly on the PC, exported through the D-Sub 9-pin connector and acquired directly in the *ManuWare* software. This will avoid the use of an external acquisition system (as, for example, the acquisition module *NI-9205*, referred at section 5.2.2).

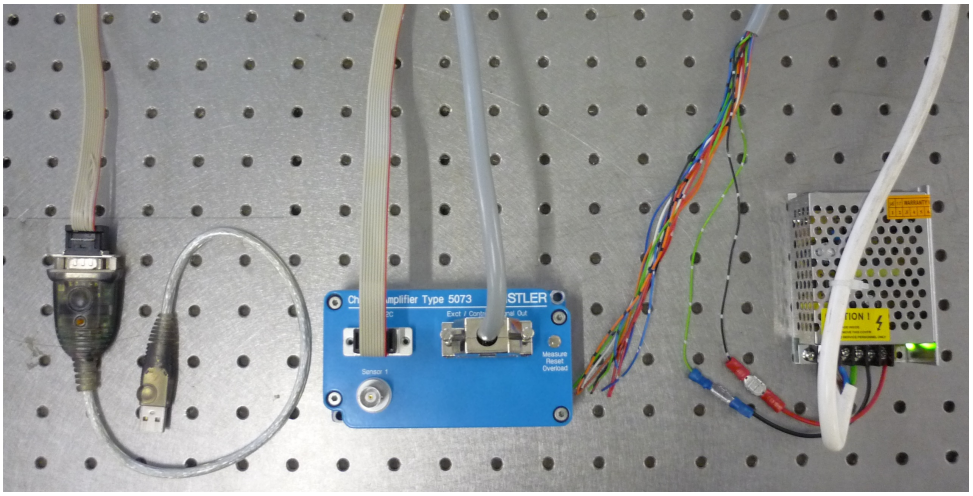


Figure 5.7: Cables connected to the D-Sub 9-pin and D-Sub 15-pin connectors.

To visualize the acquired signal in the PC, the amplifier was connected to a voltage converter (230VAC to 24VDC 1A 25W) via the pins 9 and 11 of the D-Sub 15-pin cable. The necessary cable connections is presented on the figure 5.7. However, it should be noted that the USB cable and the NBC connector need yet to be connected to the USB port of the PC and to the force sensor's cable, respectively. The *ManuWare* software, the drivers for the charge amplifier and for the USB to RS-232C (D-Sub 09 male) adapter were installed. The *ManuWare* software was executed, the device added, the channel parameters setted according to the force sensor utilized, and the measurement started. The figure 5.8 shows an example of the *ManuWare* software window during a measurement.

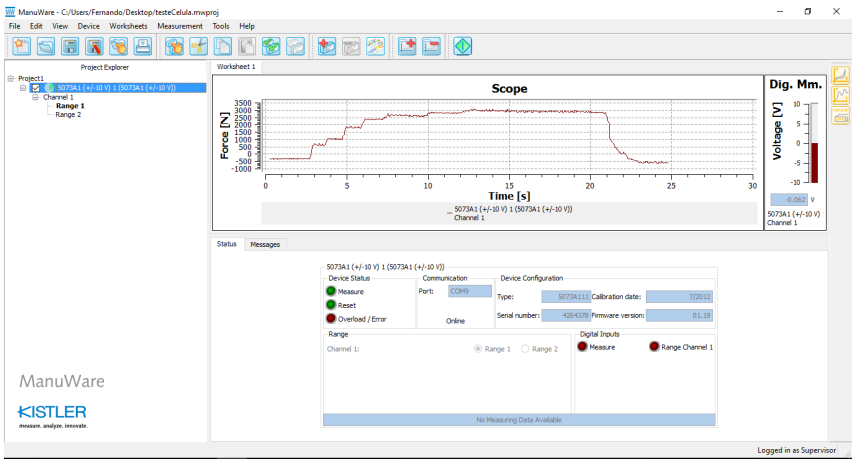


Figure 5.8: *ManuWare* software window's aspect during a regular measurement.

5.2.2 Vibration acquisition equipment

The bearings vibration can be acquired through the use of accelerometers, that can be fixed to the support, axle-bearing box or directly to the outer race of the bearing.

The equipment in use for the acquisition of the accelerometer output data was produced by *National Instruments*. The *NI cDAQ-9188* is a chassis designed for distributed sensor measurement systems, with 8 slots that allow the connection of signal acquisition's modules. The device can be connected to a PC through an ethernet cable. There are three signal acquisition modules available:

- *NI-9234*, a module with 4 analog input channels (BNC connector) and a acquisition rate of 51.2 kS/s per channel. It can measure signals from integrated electronic piezoelectric (IEPE) and non-IEPE sensors such as accelerometers, tachometers, and proximity probes.

- *NI-9211*, a module with 4 thermocouple channels (10 detachable screw-terminal connectors). It is specially built to measure signals from thermocouples.

- *NI-9205*, a module with 32 channels (36 detachable spring-terminal connectors). It can measure up to 32 single-ended or 16 differential analogue inputs, and has 4 digital channels.

We will utilize only the *NI-9234* module. The accelerometer's cable was connected to the model's BNC connector, and the *NI cDAQ-9188* was connected to the PC's ethernet port through an ethernet cable. The related software and drivers were previously installed.

Alternatively to the use of National Instrument's *LabVIEW* software, the *MATLAB* software from MathWorks was used. To assure its communication with the intended equipment, the software *NI MAX* (National Instruments Measurement & Automation Explorer) and *Data Acquisition Toolbox* were installed. *NI MAX* is a software created by National Instruments,

enabling the detection of the device and it's communication with the computer, and the *Data Acquisition Toolbox - NI-DAQmx* is a software created by MathWorks and provides support for the *MATLAB*'s acquisition of data from National Intrument's devices.

5.2.3 Data acquisition

MATLAB code

Two scripts for the environment *Matlab* were written to enable the collection of vibration data. One is for non-continuous data acquisition and the other for continuous data acquisition. A brief description of the principal lines of code will be made in table 5.2, table 5.4 and table 5.3, easing future changes or adaptation of the code (both scripts, *signalFFTnoncontinuous.m* and *signalFFTcontinuous.m*, can be found in the appendix section). In table 5.2 we can find some general configurations that are common to both cases.

Table 5.2: Parameter setting that is common to both scripts

Line	Variable	Code
6	slot	Definition of the slot's number to where the module in use is connected.
7	channel	Definition of the module's channel in use.
8	mtype	Definition of type of measurement (the accepted values are 'Voltage', 'Accelerometer', 'IEPE' and 'Microphone'). If accelerometer or IEPE is chosen, the sensitivity must be set.
10	rate	Definition of the sampling rate, in scans/second.
11	sensitivity	Definition of the sensor's sensitivity (for accelerometer and IEPE measurement type);
12	excicurr	Value of excitation current (for accelerometer and IEPE measurement type), in volts;

Continuous acquisition For the continuous acquisition of the signal, the acquisition is made in background, i.e. other code may be executed while *MATLAB* acquires value. A amount of second or scans must be defined so that, at each time that same amount is acquired, an action is executed. *MATLAB* requires this action to be described in a different file. The file *functionPlotData.m* was created, and its code can be also find in the appendix.

Non-Continuous acquisition For the non-continuous acquisition of the signal, the acquisition duration (number of scans or number of seconds) must be defined. The acquisition is made in foreground, i.e. *MATLAB* waits for the acquisition of the measured values to execute the rest of the code. See table 5.4 for more details.

5.3 Signal Analysis

One of the simplest ways of representing the acquired data is to plot the signal in the time-domain. Although it allows a quick perception of the general signal aspect, it doesn't account for many of its properties. So, three tools were developed for the analysis of the signal. The calculation of the most common statistical parameters in bearing analysis, the representation

Table 5.3: Continuous acquisition details

Line	Variable	Code
14	width	Maximum width (in seconds) of the time-domain representation of the signal (the frequency-domain representation will be made based on the equivalent number of scans).
16	secref	Definition of the amount of time between each execution of the pretended action (if secref is defined, scaref should remain undefined).
18	scaref	Definition of the amount of scans between each execution of the pretended action (if scaref is defined, secref should remain undefined).
31	d	List of the identified acquisition equipment connected to the computer.
34	s	Creation of the session (necessary step for the definition of the measurement properties and the acquisition of the measured values). For National Instruments acquisition equipment, 'ni' should be used. The instruction to a continuous acquisition is made in line 36, and the acquisition rate is defined in line 47. In line 75 it is defined what action (or, to be more specific, the function file with the action) to be performed at each given amount of time, defined in line 57 (if an amount of scans id defined) or in line 61 (if an amount of seconds is defined). If no amount is specified, the default value of a tenth of a second will be used.
66	ch	Definition of the channel from which values will be acquired (and the associated session, acquisition device, channel and measurement type). The attribution of the channel's excitation current is done on line 50 (for IEPE's measurements) or 54 (for accelerometer's measurements), and its sensitivity is defined in line 53 for accelerometer's measurements. For IEPE's measurements, the sensibility is be applied manually dividing the measurements by its value, in line 64.
75	lh	Definition of a *listener*, i.e., the action (or, to be more specific, the function file with the action) to be performed at each given amount of time, defined in line 57 (if an amount of scans id defined) or in line 61 (if an amount of seconds is defined). If no amount is specified, the default value of a tenth of a second will be used.
93	-	Instruction to initiate the measurement. Whenever the *DataAvailable* event is fired (at each amount of time defined at line 57 or 61) the *listener* in line 75 will execute the action in the related function file.

of the signal in the frequency domain based on the Fourier analysis, and the calculation of the spectral kurtosis.

In section 4.3.1 the concept and principal advantage of this techniques were already described. In the present section, we will study this techniques with more detail, and integrate them into the *MATLAB*'s acquisition code.

Table 5.4: Non-continuous acquisition details

Line	Variable	Code
16	nscans	Definition of acquisition's duration, through the total number of scans (if nscans is defined, duration should remain undefined).
17	duration	Definition of acquisition's duration, through a time interval, in seconds (if duration is defined, nscans should remain undefined).
29	d	List of the identified acquisition equipment connected to the computer.
31	s	Creation of the session (necessary step for the definition of the measurement properties and the acquisition of the measured values). For National Instruments acquisition equipment, 'ni' should be used. The attribution of the session's acquisition rate is done in line 41, and it's duration is done in line 42 (if the number of scans is defined) or in line 44 (if the number os seconds is defined).
48	ch	Definition of the channel from which values will be acquired (and the associated session, acquisition device, channel and measurement type). The attribution of the channel's excitation current is done on line 50 (for IEPE's measurements) or 54 (for accelerometer's measurements), and its sensitivity is defined in line 53 for accelerometer's measurements. For IEPE's measurements, the sensibility is be applied manually dividing the measurements by its value, in line 64.
59	result, t	The measurement values are obtained (data acquired and respective time stamps). As this is a foreground operation, the rest of the code is only executed after the acquisition of the measurement values.

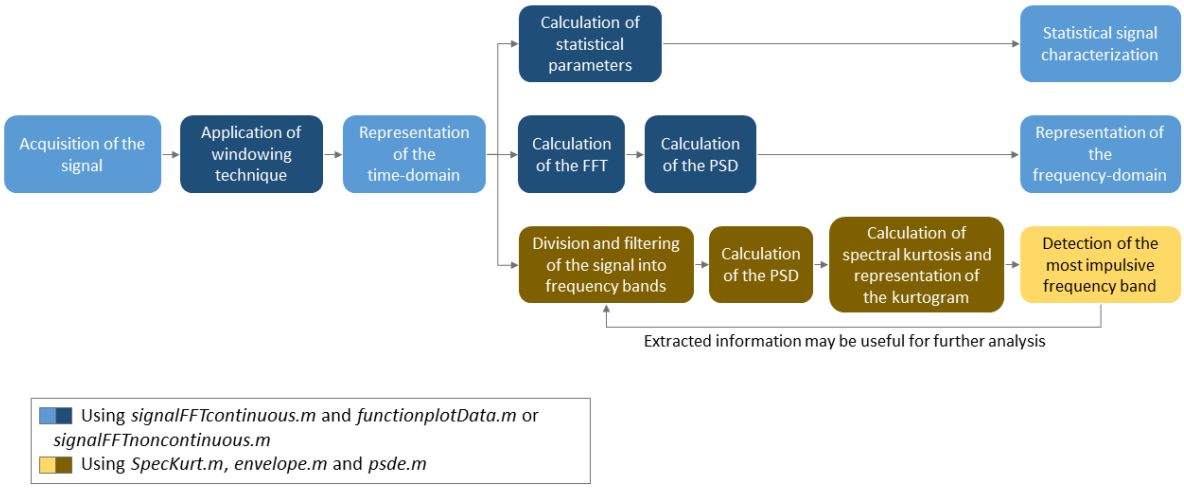


Figure 5.9: Flowchart indicating the main steps of the signal analysis process adopted.

5.3.1 Statistical Parameters

Theory

As we have already seen in 4.3.1, the calculation of statistical parameters is a very simple procedure that, although should not be used alone, can be a good indicative of signal faults.

A number of simple signal metrics based on the time domain waveform are used in mechanical fault detection. Most of statistical parameters of a signal can be obtained from the probability density function. The most common statistical parameters used for bearing monitoring is root mean square (RMS), crest factor, skewness and kurtosis.

The Root mean square (RMS) is the normalized second central moment of the signal. It is a statistical measure of the magnitude of a varying quantity, that represents the overall vibration energy. RMS is not sensitive to sudden short duration and isolated peaks in the signal, so it does not give meaningful information to identify which component is failing, but can effective in tracking system noise [40]. According to [41], it can be defined as

$$\text{RMS} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(x(n)-\bar{x}\right)^2} \tag{5.1}$$

where

$$\bar{x} = \frac{1}{N}\sum_{n=1}^{N} x(n) \tag{5.2}$$

Crest factor is the ratio of the peak value of the input signal to the RMS value, and is typically used on the raw vibration signal. It detects faults through changes in the signal pattern due to impulsive vibration sources [40] and is useful to detect discrete impulses above the background signal which do not occur frequently enough or have sufficient duration to significantly increase the RMS level. A larger value of the crest factor will correspond to larger impulses, as the peak value of the impulsive signal is significatly changed while the RMS level of the overall signal is only slightly changed [40]. It can be defined as

$$\text{Crest factor} = \frac{\max\left(x(n)\right)}{\sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(x(n)\right)^2}} \tag{5.3}$$

Skewness is the third centred moment of the probability density function. It has a value of zero for symmetrical functions and a large one for asymmetrical functions. It is usually normalized by being divided by the appropriate power (of 3) of the standard deviation [42]. It can be defined as

$$\text{Skewness} = \frac{\frac{1}{N}\sum_{n=1}^{N}\left(x(n)-\bar{x}\right)^3}{\left(\sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(x(n)-\bar{x}\right)^2}\right)^3} \tag{5.4}$$

Kurtosis is the fourth moment of the probability density function. It is a non-dimensional quantity used to detect the presence of significant peaks in the time-domain of the vibration signal. It has a large value for sharper peaks in the signal (specially in impulsive signals) at the signal is raised to the fourth power. This amplifies effectively the isolated peaks in the signal. As it happen with skewness, kurtosis is normalized being divided by the appropriate power of the standard deviation [42]. It can be defined as

$$\text{Kurtosis} = \frac{\frac{1}{N}\sum_{n=1}^{N}\left(x(n)-\bar{x}\right)^4}{\left[\frac{1}{N}\sum_{n=1}^{N}\left(x(n)-\bar{x}\right)^2\right]^2} \tag{5.5}$$

MATLAB code

Although *MATLAB* has already intrinsic functions for the calculus of various statistical parameters, it was decided to used the formulation presented above.

For continuous acquisition, the code for the calculation of statistical parameters can be found in *functionplotData.m*, from line 56 to 62 or from line 82 to 88. For non-continuous acquisition, the code can be found in the *signalFFTnoncontinuous.m* file, from line 108 to 122. The code for both files can be found in the appendix section.

5.3.2 Fast Fourier transform

Theory

The basic concept of Fourier analysis consists in the demodulation of a signal in a summation of sinusoidal components (the following deduction is made based on [42]). For any given signal $g(t)$ with a integer number, n, of periods, T, can be expressed as

$$g(t) = g(t + nT) \tag{5.6}$$

or decomposed as a sum of sines and cosines,

$$g(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos(k\omega_0 t) + \sum_{k=1}^{\infty} b_k \sin(k\omega_0 t) \tag{5.7}$$

where ω_0 is the fundamental angular frequency in $\mathrm{rad\,s^{-1}}$. a_k and b_k are obtained through

$$a_k = \frac{2}{T} \int_{-T/2}^{T/2} g(t) \cos(k\omega_0 t) \mathrm{d}t \tag{5.8}$$

$$b_k = \frac{2}{T} \int_{-T/2}^{T/2} g(t) \sin(k\omega_0 t) \mathrm{d}t \tag{5.9}$$

Alternatively, equation 5.7 can be written as a sum of rotating vector as

$$g(t) = \sum_{k=-\infty}^{\infty} A_k \cos(j\omega_k t) \tag{5.10}$$

where

$$A_k = \frac{1}{T} \int_{-T/2}^{T/2} g(t) e^{-j\omega_k t} \mathrm{d}t \tag{5.11}$$

The Fourier transform can be derived from the Fourier series, allowing the periodic time to tend to infinity and at the same time removing the division by T because transients have finite energy rather than finite power. Equations 5.10 and 5.11 then become

$$G(f) = \int_{-\infty}^{\infty} g(t) e^{-j2\pi f t} \mathrm{d}t \tag{5.12}$$

$$g(t) = \int_{-\infty}^{\infty} G(f) e^{-j2\pi f t} \mathrm{d}f \tag{5.13}$$

respectively, where angular frequency ω_k has been replaced by the continuous frequency function f expressed in Hz. These equations are known as forward and inverse Fourier integral transforms, respectively.

For digitalized signal the Discrete Fourier transform should be used. When the signal is digitalized, it becomes discretely sampled. As the record length is finite, it leads to the same situation as with the Fourier series in which the spectrum is discrete and the time record implicitly periodic [42]. The continuous infinite integrals of the Fourier transform become finite sums, expressed as

$$G(k) = \frac{1}{N} \sum_{n=0}^{N-1} g(n) e^{-j2\pi kn/N} \tag{5.14}$$

where

$$g(n) = \sum_{k=0}^{N-1} G(k) e^{j2\pi kn/N} \tag{5.15}$$

In equation 5.14 the division by the length of the signal, N, corrects the scaling of the Fourier series components.

Alternatively, the Discrete Fourier transform it can be understood as the matrix multiplication

$$\mathbf{G}_k = \frac{1}{N}\mathbf{W}_{kn}\mathbf{g}_n \tag{5.16}$$

where $\mathbf{G}_k$ represents the vector of N frequency components, $\mathbf{g}_n$ represents the N time samples and $\mathbf{W}_{kn}$ a square matrix of unit vector $\exp(-j2\pi kn/N)$ with angular orientation depending on the frequency index k (rows) and time sample index n (columns). This is illustrated on figure 5.10.

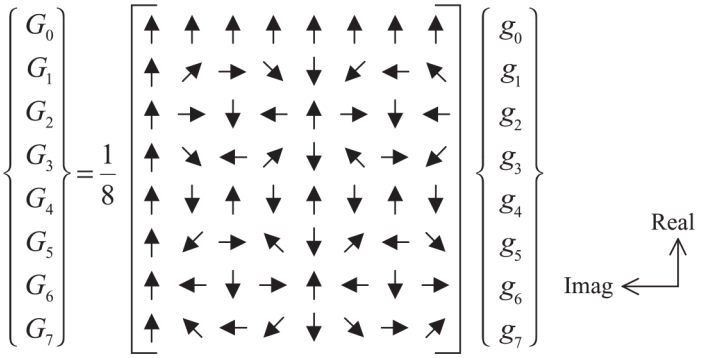


Figure 5.10: Matrix representation of the Discrete Fourier transform [42].

The Fast Fourier transform is an efficient algorithm and one of the simpler ways for an optimized implementation of Discrete Fourier transform to a signal. The matrix $\mathbf{W}_{kn}$, from equation 5.16, is constructed assuming that N is a power of 2. The modified version of the matrix $\mathbf{W}_{kn}$ is then factorized into $\log_2 N$ matrices. So, for each multiplication by each matrix it is only necessary N complex operations, unlike the N^2 operations required for the multiplication by a conventional matrix [42]. The modified version of $\mathbf{W}_{kn}$ is arranged in a different order of the original matrix. It is represented in figure 5.11, where is evident that this new matrix can be factorized into three matrices, obtaining only two non-zero elements per row.

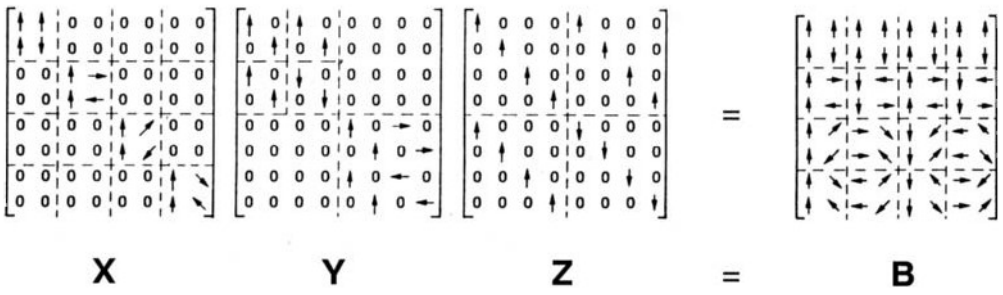


Figure 5.11: Matrix representation of the Fast Fourier transform [42].

Power Spectral Density There are many possible ways of visually represent the fft results. The Power Spectral Density is one of the possible representation techniques, and is useful when comparing to distinct signal.
The power of each frequency component can be calculated through

$$P_x(f) = X(f)X^*(f) \tag{5.17}$$

where $X(f)$ is the frequency domain representation of the signal $x(t)$. As what is intended is the power spectral density, the result should be scaled, diving its value by the bandwidth of the signal

$$PSD_x(f) = \frac{X(f)X^*(f)}{S_r l_x} \tag{5.18}$$

where S_r represents the sampling rate, and l_x correspond to the length of the signal (number of elements of the signal considered for the calculation of the Fast Fourier transform).

Windowing technique The Fourier transform assumes that the finite data collected is a continuous spectrum of a one period of a periodic signal. However, the measured signal doesn't correspond to an integer number of periods since it is a combination of sinusoidal waves, at different frequencies. When the signal collected correspond to a truncated waveform, discontinuities are introduced into the calculation, showing up in the transformed signal as high-frequency components (that aren't present in the original signal). These frequencies are often higher than the Nyquist frequency, and give the sensation that the energy at one frequency leaks into other frequencies, as fine spectral lines are spread into the wider signals. This phenomenon is thus known as spectral leakage[43].

This effect can be minimized using a technique called windowing, which reduces the amplitude of the discontinuities at the boundaries of each finite sequence of acquired data. Windowing consists of multiplying the time record by a finite-length window with an amplitude that varies smoothly and gradually toward zero at the edges, thus making the endpoints of the waveform meet. A continuous waveform without any sharp transitions is achieved.

There are various different Windowing techniques that can be applied, depending of the signal acquired. If a window from one of the Windowing tecniques is plotted, a continuous spectrum with a main lobe (in the center) and several side lobes (at each side) are obtained. The main lobe is centered at each frequency component of the time-domain signal and the side lobes approach zero in both ends. Each technique varies the height of the side lobes, which are related to how much leakage exists in the measured FFT. This will affect the bandwidth of the major lobe. So, a equilibrium between the height of the side lobes and of the main lobe must be achieved[43]. The figure 5.12 is a representation of a table with different windowing techniques for different signal types.

Signal Content	Window
Sine wave or combination of sine waves	Hann
Sine wave (amplitude accuracy is important)	Flat Top
Narrowband random signal (vibration data)	Hann
Broadband random (white noise)	Uniform
Closely spaced sine waves	Uniform, Hamming
Excitation signals (hammer blow)	Force
Response signals	Exponential
Unknown content	Hann
Sine wave or combination of sine waves	Hann
Sine wave (amplitude accuracy is important)	Flat Top
Narrowband random signal (vibration data)	Hann
Broadband random (white noise)	Uniform
Two tones with frequencies close but amplitudes very different	Kaiser-Bessel
Two tones with frequencies close and almost equal amplitudes	Uniform
Accurate single tone amplitude measurements	Flat Top

Figure 5.12: Diverse Windowing techniques that can be used for different types of signals [43].

MATLAB code

MATLAB as an integrated function for the calculation of the Fast Fourier transform, which simplifies the process using this type of transformation. However, some attention must be given to the following details:

- The utilization of a number of points that is a power of two (although the *fft* function of *MATLAB* can use of any number of points).
- The representation of the frequency domain in a power spectrum density plot.
- The utilization of a windowing technique.

The quantity of points utilized in this calculation is always a power of two. If the number of point is superior or inferior to this number, the closest, inferior number will be used. For a non-continuous acquisition, the total number of points used corresponds to the number of point acquired for the specified duration of measurement. For a continuous acquisition, at each update of the time-domain plot, the total number of points used correspond to the number of points that are being displayed at the time-domain plot.

The windowing technique choosen was the hanning window, as it gives a satisfactory response in 95 percent of cases, and has a good frequency resolution and reduced spectral leakage [43].

A *MATLAB* code was written for the calculation of the Power Spectral Density, and integrated into the script for data acquisition. For non-continuous acquisition, the code is located into the main script file (*signalFFTnoncontinuous.m*), but for the continuous acquisition it had to be inserted into the function file (*functionPlotData.m*), as seen in section 5.2.3. A short description of the principal lines of code of both files can be found at table 5.5.

5.3.3 Spectral Kurtosis and Kurtogram

Theory

Spectral Kurtosis is a useful tool to determine which frequency band contain a signal of maximum impulsivity. It can be used to diagnose machine faults that give rise to a series of impulse responses. For rolling bearings, a response of this kind, $Y(t)$ can be modelled as a signal excited by impulses X at times τ_k

$$Y(t) = \sum_k g(t - \tau_k)X(\tau_k) \tag{5.19}$$

The Short time Fourier is obtained by shifting a time window along the record may be represented in terms of the amplitude envelope function $H(t, f)$, but its square value represent the power spectrum values at each position. The average of all these short time power spectra may be obtained by the average of all these short time power spectra [42].

The kurtosis for each frequency, f, can be calculated by taking the fourth power of $H(t, f)$ at each time and averaging its value along the record, and then normalizing it by the square of the mean square values, as shown in the following equation

$$K(f) = \frac{H^4(t, f)}{[H^2(t, f)]^2} \tag{5.20}$$

The success of the results depends strongly on the values chosen for the time window used for the calculation of the successive Short time Fourier transforms. Although the denominator of the equation 5.20 is independent of the window length chosen, the same can not be said about the numerator. For a maximum value of kurtosis, the window must be shorter than the spacing between the pulses but longer than the individual pulses [42].

The spectral kurtosis will have a large value for frequency bands where the fault signal is dominant and small where the spectrum is dominated by stationary signal. The kurtogram is

Table 5.5: Principal lines of code for the calculation and representation of the frequency domain.

Line		Code
Non-cont.	Cont.	
67,70	40, 42 or 56, 58	Calculation of the closest, inferior power of two of the number of points acquired (for non-continuous signal) or of the number of point considered for the time-domain plot.
76	45 or 61	Application of a Windowing (Hann) to the time signal.
77	46 or 62	Application of the Fast Fourier transform to the signal. As the windowing technique applied to the signal was a hanning window, the result must be multiplied by a factor of 2.0 for amplitude correction.
78	47 or 63	Calculation of the power spectral density of the signal.
79	48 or 64	Selection of the positive frequency components. The Fourier transformation outputs a result that considers both a positive and negative frequency, and for real signals, the result is about the zero frequency.
81	49 or 65	Calculation of the corresponding frequency, for the considered number of points.
82	50 or 66	Selection of half of the frequency, that corresponds to "half" of the selected power spectrum. It is coherent to the fact of the Discrete Fourier transform only calculates spectral components up to half of the sampling frequency.
94	52 or 68	Plot of the Power Spectral Density

a map representation of the spectral kurtosis, helping in the selection of the optimum center frequency and bandwith combination [44].

The kurtogram is a visual representation of the value of the kurtosis for different frequency bands of the signal. It is divided in various *levels*, each one containing a certain amount of frequency bands (higher levels correspond to a higer number of considered frequency bands). For each frequency band, the value of its kurtosis is calculated. The kurtogram allows thus for a quick identification of the signal's frequency band that is more impulsive (higher value of kurtosis), usually containing the frequency of the faulty signal.

MATLAB code

The code used for the calculation of the kurtogram was found in [44]. Only a few alterations were made, enabling the automatic usage of the values obtained from the scripts written for data acquisition.

For the calculation of the Kurtogram, one script (*SpecKurt.m*) and two functions are used (*envelope.m* and *psde.m*). Although the two functions play a part in the calculation of the spectrum, they can also be used alone to analyse the acquired data. *envelope.m* allows the calculation of the envelope of the signal for the defined frequency limits. *psde.m* provides the calculation of the Power Spectrum.

The code *psde.m* is similar to the one developed for the calculation of the power spectral density (the most notorious difference being the fact that the last one divides the power spectrum by the frequency band, generating a spectral density). For the calculation of the spectrum, *psde.m* uses various windows that overlap, calculating various spectrums. The sum of all these spectrums is the normalized, resulting in the spectrum of the original input signal. This method differs from the one written for the calculation of the power spectral density, which only considers a single window (containing the whole signal).

Contrary to what was made for the previous sections of *Statistical Parameters* and *Fast Fourier transform*, there is no need for a table that explains the principal lines of code, since the code of each script/function (which is presented in the appendix section) as already some annotations explaining the process. However, a brief description will be made, explaining the overall procedure of each script/function.

SpecKurt.m A range of eight octave was chosen for the construction of the kurtogram. For each octave, the signal is divided into the corresponding number of bands. For each octave, the functions *envelope.m* and *psde.m* are used. The first function calculates the envelope of the band, and the second uses it to calculate its power spectrum. The kurtosis of each band's spectrum is calculated, and stored in a map, the kurtogram. For each level of the Kurtogram, there is a kurtosis value for each level's band.

envelope.m The signal is converted to its frequency domain through the Fast Fourier transform. The resulting signal that is within the band's limit frequencies is copied into the base band (Heterodyne operation) and the value of zero is assigned to the remaining values (Fourier filtering and Hilbert transform). The inverse Fast Fourier transform is then calculated, and the absolute value of the resulting vector corresponds to the output of the function.

psde.m The signal is divided into different time windows that overlap each other. For each window, the Hanning window technique is applied and the linear trend is then removed. The Fast Fourier transform is then calculated and each value is raised to the power of 2. Then, the various window's vector are summed altogether and divided by a scale factor, resulting in single, normalized, vector. It represents the band's frequency-domain.

5.4 Validation of results

5.4.1 Acquisition system

To validate the results obtained through the equipment, a known signal was generated and compared to the measured results.

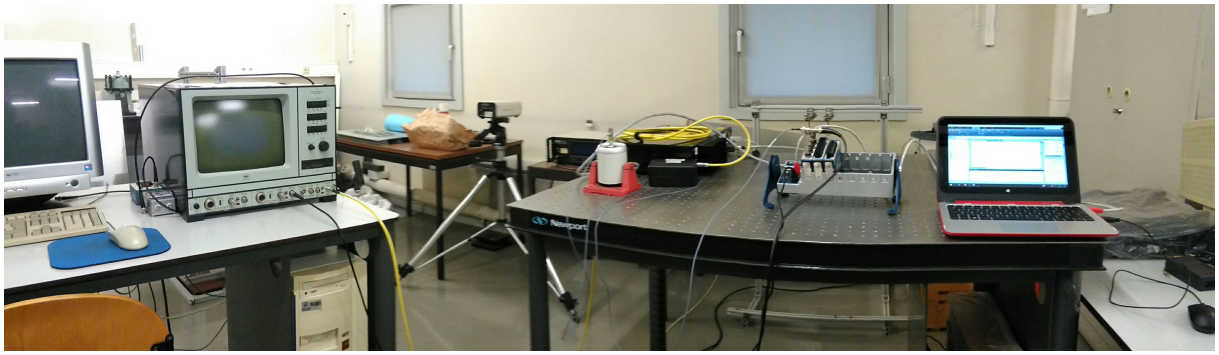
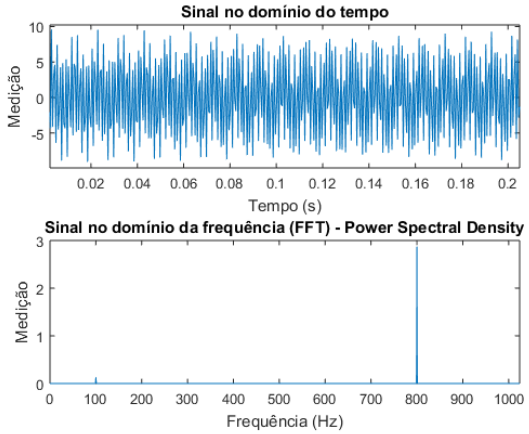
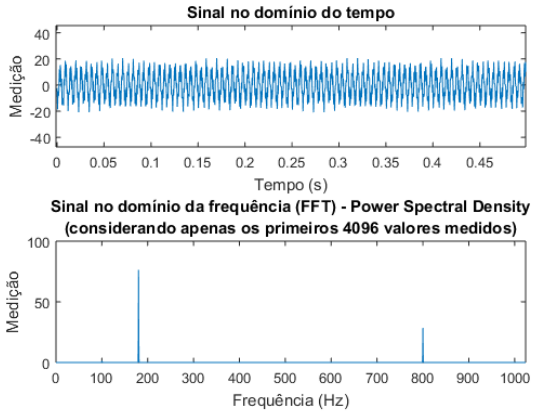


Figure 5.13: Montage used for the validation of the acquisition's system results.

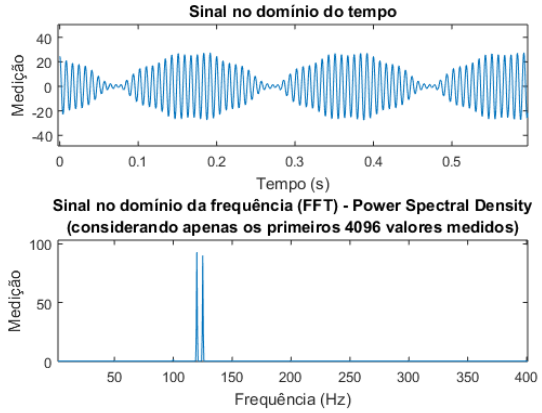
The signal was generated through a signal analyser, *Brüel & Kjær, Signal Analyzer Unit type 2035*, connected to a electrodynamic shaker, *Ling Dynamic System, Permanent Magnet Shacker V201*, through a power amplifier, *Ling Dynamic System, Power Amplifier PA25E*. The accelerometer used in conjunction with the acquisition system, *Dytran, Accelerometer 3136A*, was fixed on the electrodynamic shaker, and the signal was measured with the *National Intruments* acquisition equipment.
The accelerometer was then connected directly to the signal analyser, and the same signal was acquired and visualized, and compared to the data previously acquired.



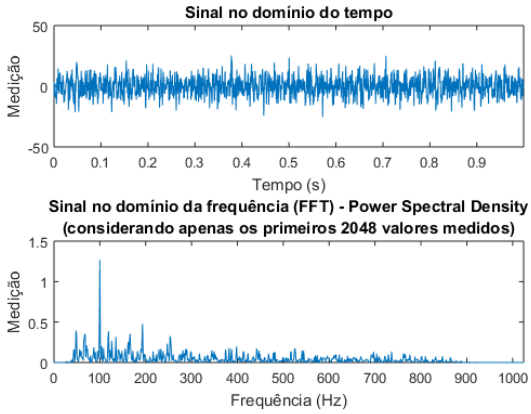
(a) A sinusoid generated at 800 Hz.



(b) A sum of two sinusoids, generated at 180 and 800 Hz.



(c) A sum of two sinusoids, generated at 120 and 125 Hz.



(d) Noise signal

Figure 5.14: Results of different acquisitions.

Various signal were generated, and the results recorded. We will display the measured result of four different acquisitions:

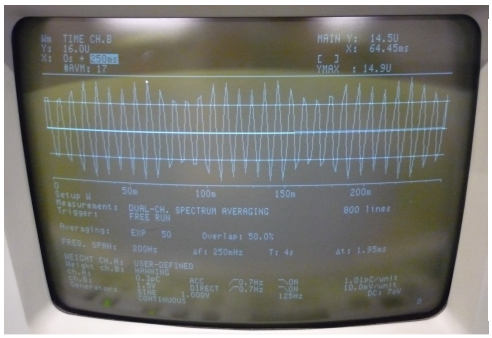
- The acquisition of a sinusoid generated at 800 Hz (displayed in figure 5.14a)
- The acquisition of a signal composed by two sinusoids, one generated at 180 Hz and the other at 800 Hz (displayed in figure 5.14b).
- The acquisition of a beat frequency, a signal composed by two sinusoids generated at close frequencies. One of them was generated at 120 Hz and the other at 125 Hz (displayed in figure 5.14c).

- The acquisition of noise, a random generated signal (displayed in figure 5.14d).

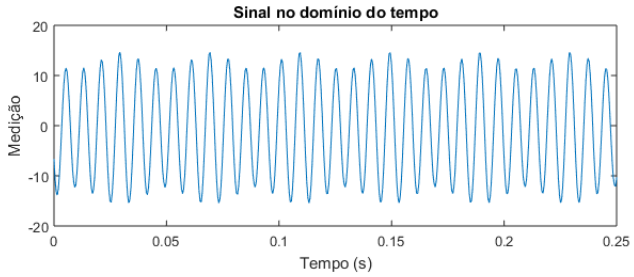
Analysing the overall aspect of the time-domain representation, and comparing the frequency-domain representation to each signal's frequency, it can be seen that the result obtained are similar to what was expected. It should be noted, however, that the use of the frequency-domain representation is simply to surpass the manual calculation of each signal's frequency, and not to validate the results obtained through it. The validation of the frequency-domain results can be found in 5.4.2

5.4.2 Fast Fourier Transform

In an attempt to test the values of the power spectrum density, we will compare the results obtained to the ones visualised in the signal analyser.

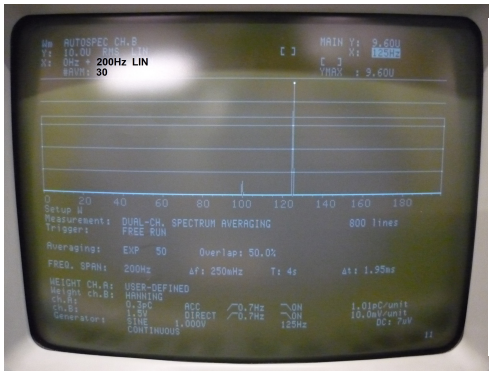


(a) Acquired in the signal analyser.

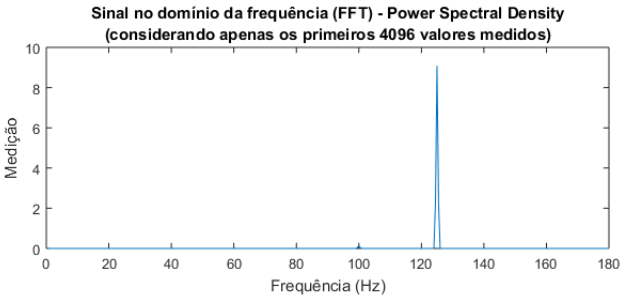


(b) Acquired with the National Instruments equipment.

Figure 5.15: Time domain representation of the signal.



(a) Acquired in the signal analyser.



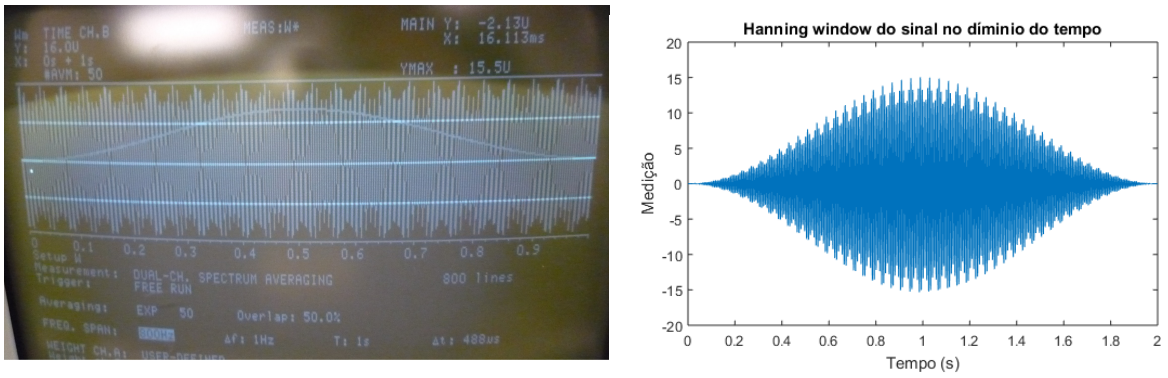
(b) Acquired with the National Instruments equipment.

Figure 5.16: Frequency domain representation of the signal.

A signal was generated, a sinusoid with a frequency of 125 Hz. The power density spectrum was generated, both on the signal analyser and through the used of the *National Instruments* equipment. The results can be visualized at figures 5.16, 5.15 and 5.17.

The time domain representation and windowing techniques are similar in both side-by-side images. The frequency-domain representation may appear different, because of the existence of a smaller peak, in the signal acquired by the analyser. This peak at 100 Hz is an interference from a harmonic of the frequency of the power supply, and shouldn't be considered since it isn't created by the signal.

It can be concluded that the acquisition system assembled is correctly acquiring the data and representing both the time-domain and the power spectral density.



(a) Applied by the signal analyser.

(b) Applied in MATLAB.

Figure 5.17: Windowing technique (Hanning) applied to the signal.

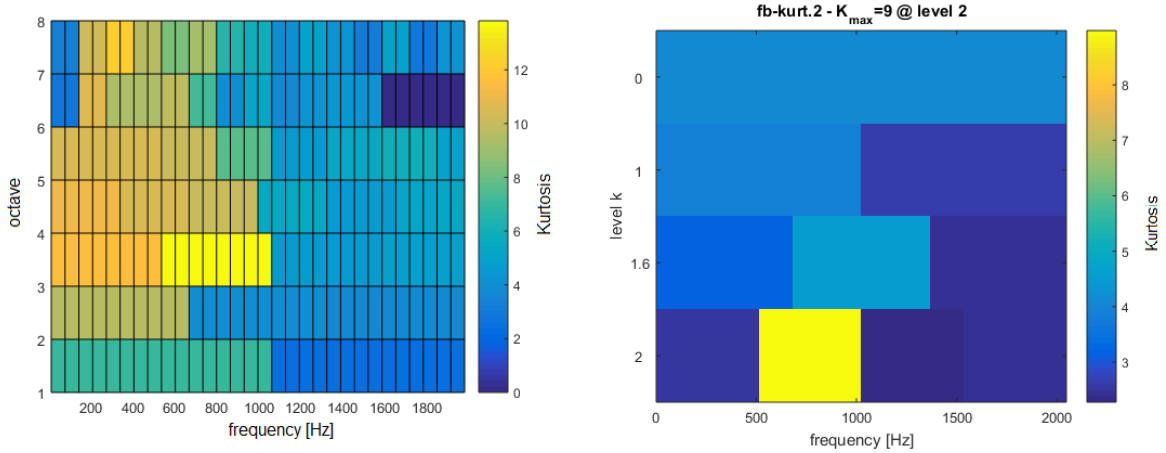
5.4.3 Spectral Kurtosis and Kurtogram

To validate the results obtained through the Kurtogram (*SpecKurt.m*), another *MATLAB* function, found at [45], was used. Both functions were applied to various measurements and the results compared. At figure 5.18, we can compare the results using the two methods. It is clear that both find a maximum value of Kurtosis for a frequency range of 500-1000 Hz.

To test *envelope.m*, various envelopes were created and compared to the original signal. In figure 5.19, as expected, the envelope covering a large band of the signal (0-2048 Hz) is similar to the original signal (sampling rate of 4096 Hz), whereas envelopes covering smaller bandwidths tend to diverge from the original signal.

To test *psde.m*, the output of the function was compared to the Power spectral density obtained automatically when the signal was acquired. The results are similar (the power spectrum obtained through *psde.m* was scaled to match the power spectral density values), although they differ to some level due to the differences of the both methods (already discussed in section 5.3.3).

To test *SpecKurt.m*, *envelope.m* and *psde.m*, the kurtogram of a measured signal was calculated. Though its analysis, the envelopes of the highest kurtosis's frequency band were produced, and then their power spectrum calculated. In figure 5.21, the results are as expected, as it can be seen higher peaks for the envelopes that correspond to higher kurtosis value.



(a) Using (*SpecKurt.m*).

(b) Using the function found in [45].

Figure 5.18: Resulting Kurtogram

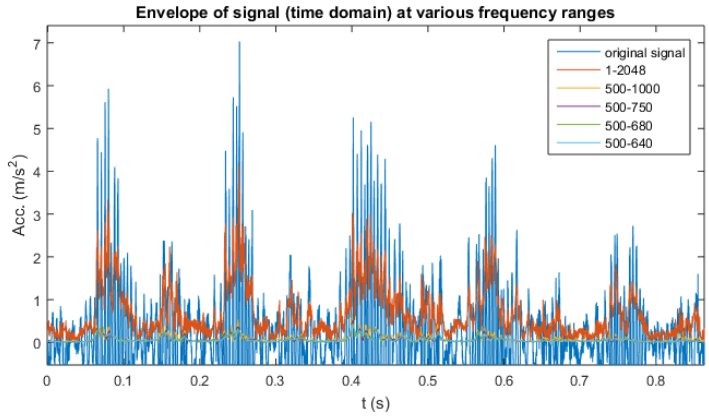


Figure 5.19: Comparison of the original measured signal to various envelopes.

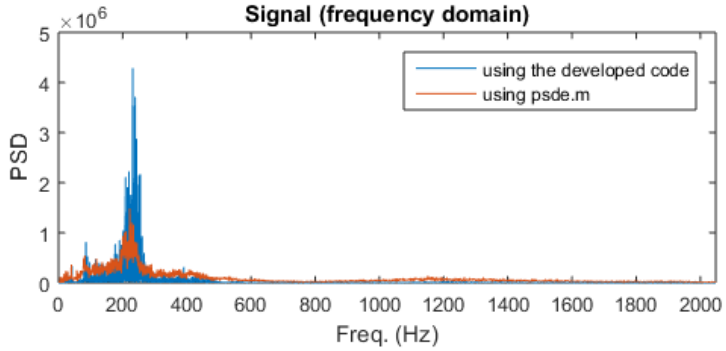
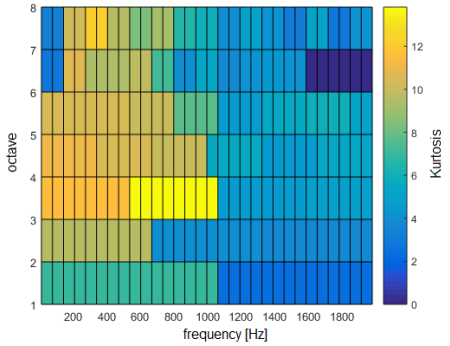
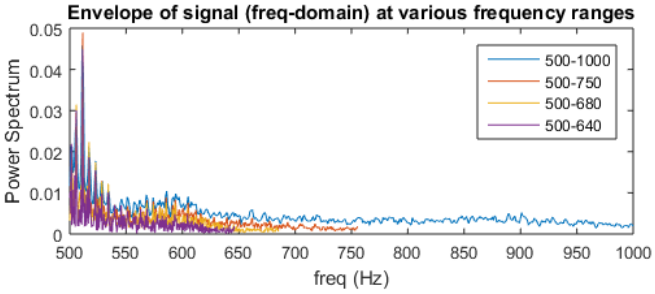


Figure 5.20: Comparison of the signal's power spectral density and the result obtained using *psde.m*.



(a) Kurtogram.



(b) Plot of the power spectrum of several envelopes.

Figure 5.21: Power spectrum of several envelopes, which bandwidths were chosen according to the kurtogram.

5.5 Previous calculations

For the use of the test bench, it is necessary to know the equivalent load to apply to the bearing that mimics the real application load. It is important to have a perception of the bearings characteristics frequencies when analysing the measured signal, since these frequencies can be directly link a failure to its location. The procedure followed for the calculation of these values is shown below.

5.5.1 Calculation of dynamic bearing loads

The dynamic bearing loads can be calculated based on the value of external forces as axleload, weight of the wheelset and payload. The following procedure is given by [6] and permits the calculation of load components for a single bearing. It considers the axle journal as a beam resting on a rigid, moment-free bearings supports. Elastic deformation in the bearing, axlebox housing or bogie frame components is not considered and the moments on the bearing due to elastic deflection of shaft, axlebox housing or bogie are not taken into account (the shaft in not the only component that creates moments).

Firstly, the overall vehicle load and number of wheelsets and bearings must be known. The static axleload corresponds to the maximum load carried by the train, i.e., the value of all the passengers weight (considering overload) plus the weight of the vehicle in running order. As the wheelset weight is not given, we will estimate it through a rough estimation of the bogies weight divided by a factor of 2 (considering that half of its weight belongs to wheelsets components). The unit has a total of 12 wheelsets. The overall technical data and schematic design of the UTE 2240 unit consulted can be found in the appendices section.

Table 5.6: UTE's considered weight, in Kg

Passengers mass at overload (70 Kg/passenger)	Mass of the vehicle in running order	Wheelset mass
40.040×10^3	135.700×10^3	2.932×10^3

Note that the the powered car carries a greater load than the non-powered cars and that the concerned bearing (SKF's TBU 1639590 A) is only used in the UTE's non-powered cars. Dividing the total unit load by the number of wheelsets will permit to over-estimate the maximum load applied to these bearing's wheelsets (guaranteeing a safety factor on the calculations).

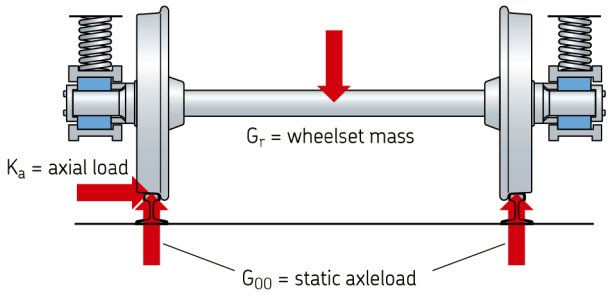


Figure 5.22: Considered acting bearing loads [6].

The static bearing load, G, can be calculated according,

$$G = \frac{G_{oo} - G_r}{2} \tag{5.21}$$

where G_{00} is the maximum static axlebox load, G_r is the weight of the wheelset. The value of G can be found in table 5.7.

Table 5.7: Values used in the calculation of maximum static axlebox load, G, in KN

G_{00}	G_r	$\boldsymbol{G}$
143.668	28.738	**57.464**

The mean radial load, K_r is based on the static bearing load, G, and should consider variations in the payload as well as any additional dynamic radial and traction forces. It can be calculated according to,

$$K_r = f_0 f_{rd} f_{tr} G \tag{5.22}$$

where f_0 is the payload factor, f_{rd} is the dynamic radial factor, and f_{tr} is the dynamic traction factor. f_0 is used to take into account variation in the static loading of the vehicle, f_r to take into account quasi-static effects (rolling and pitch) and dynamic effects (from the wheel rail contact) and f_{tr} to take into account additional radial loads caused by the drive system.

For *multiple units, passenger coaches and mass transit vehicles*, the value of f_0 can range from 0.9 to 1 and the value of f_{rd} can range from 1.1 to 1.3. As the concerned bearing vehicle in non-powered, the value of f_{tr} should be 1. The value of K_r can be found on table 5.8.

Table 5.8: Values used in the calculation of mean radial load, K_r (G and K_r in KN)

f_0	f_{rd}	f_{tr}	G	$\boldsymbol{K_r}$
0.95	1.2	1	57.464	**65.510**

The mean axial load, K_a is based on the mean axial load, G, and should consider the dynamic axial forces. It can be calculated according to,

$$K_a = f_0 f_{ad} G \tag{5.23}$$

where the payload factor, f_0, is already known and f_{ad} is the dynamic axial factor. For *multiple units, passenger coaches and mass transit vehicles, maximum speed 160 km/h*, the value of f_{ad} is 0.08 (vehicles maximum speed is 120 km/h). The value of K_a can be found on table 5.9.

Table 5.9: Values used in the calculation of mean radial load, K_a (G and K_a in KN)

f_0	f_{ad}	G	$\boldsymbol{K_a}$
0.95	0.08	57.464	**4.367**

The radial bearing load, F_r and the axial bearing load, F_a, can only be calculated after knowing the value of the mean radial load and mean axial load, respectively.

For cases where the force is acting symmetrically, either on top of the axlebox, or on both sides, there is no need for special calculations as, for example, axleboxes with link arms (non-symmetrical axleboxes). In the UTE 2240, the axle-box can be considered symmetrical (see figure 5.23).

For double-row tapered roller bearings, the radial bearing load, F_r, can be calculated according to,

$$F_r = K_r + 2 f_c K_a \tag{5.24}$$

where K_r and K_a are already known, and f_c is a factor that takes into account the bearing geometry. It can be calculated through:

$$f_c = h \frac{D_a}{l} \tag{5.25}$$

where h assumes the value of 0.1 (as the load acts near to the central plane of the bearing), D_a represents the shaft diameter and l the distance between the 2 load center (see figure 5.24).

The value of F_r can be found on table 5.10.

The value of bearing axial load, F_a, is equal to the value of mean axial load:

$$F_a = K_a \tag{5.26}$$

The values of the forces acting are synthesised on table 5.11.

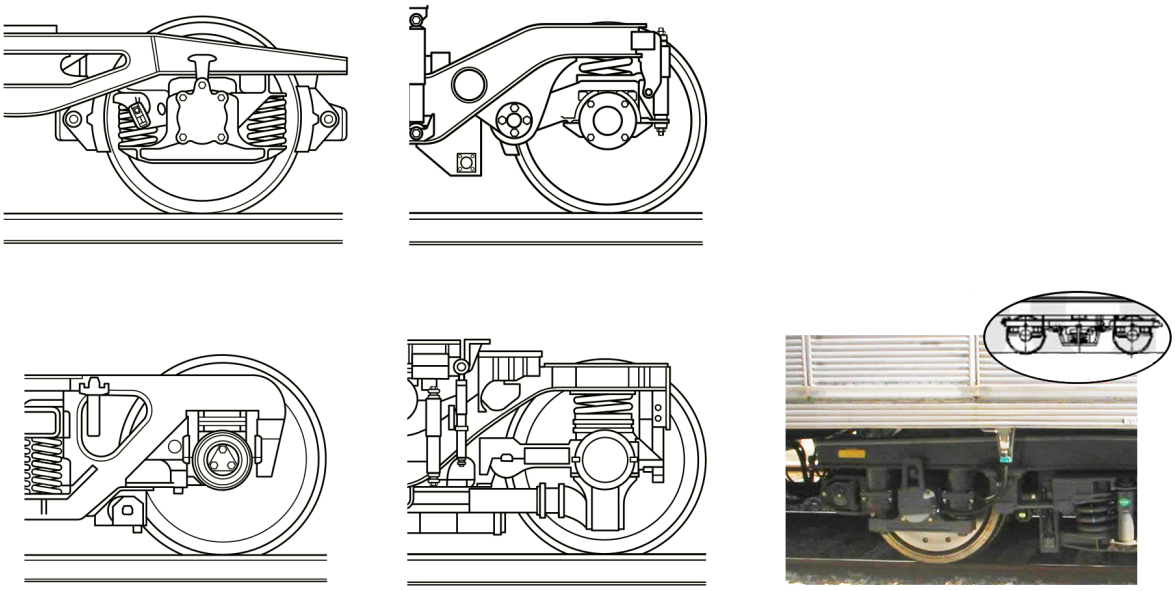


Figure 5.23: Exemple of symmetrical axleboxes designs (on the left), of non-symmetrical axleboxes designs (on the center) and a picture of the axlebox of the UTE 2240, symmetrical(on the right)

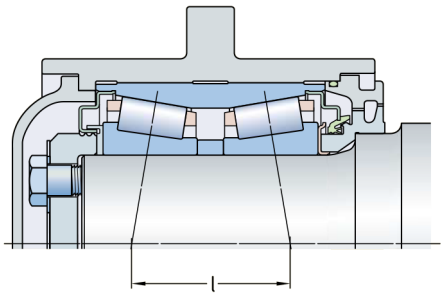


Figure 5.24: Schematic representation of value l, the distance between 2 load center of a double-row tapered roller bearing [6].

Table 5.10: Values used in the calculation of mean radial load, F_r (D_a and l in mm and K_r, K_a and F_r in KN)

h	D_a	l	f_c	K_r	K_a	$\boldsymbol{F_r}$
0.100	130.0	112.1	0.116	65.510	4.367	**66.522**

Table 5.11: Values of radial and axial bearing loads, K_r and K_a, respectively, in KN

K_r	K_a
66.552	4.367

5.5.2 Calculation of bearing characteristic frequencies

As we have already seen in chapter 2, section 2.3, the bearing characteristic frequencies can be obtained as specified in equations 2.1, 2.2, 2.3 and 2.4, rewritten below:

$$\text{BPFO} = \frac{nf_r}{2}\left\{1 - \frac{d}{D}\cos\phi\right\} \tag{5.27}$$

$$\text{BPFI} = \frac{nf_r}{2}\left\{1 + \frac{d}{D}\cos\phi\right\} \tag{5.28}$$

$$\text{BSF} = \frac{D}{2d}\left\{1 - \left(\frac{d}{D}\cos\phi\right)^2\right\} \tag{5.29}$$

$$\text{FTF} = \frac{f_r}{2}\left\{1 - \frac{d}{D}\cos\phi\right\} \tag{5.30}$$

The TBU 1639590 A has a pitch diameter of 166.7 mm, a total of 42 rollers (21 rollers/row) each with a diamter of 20.4 mm and a contact angle of 9º. For the shaft frequency, a range of values will be defined, permitting the representation of the evolution of the bearing characteristic frequencies with the shaft frequency. The upper limit will be the frequency that corresponds to the maximum velocity achieved by the train, and the lower limit to a minimum threshold, as for small values of velocity the noise levels produced by the vehicle are considerable and the collection of signal data is discouraged.

The shaft velocity is directly related to the the train velocity and can be calculated through the trains wheel diameter, D_{wheel} (see equation 5.31). The upper shaft frequency considered correspond of the UTE 2240's maximum velocity, 120 Km/h, and the lower limit to an established value of 40 Km/h.

$$f_r\,[\text{Hz}] = \left(vel\,[\text{km/h}] \times \frac{1000}{3600}\right)\frac{1}{\pi D_{wheel}} \tag{5.31}$$

The following values were considered: 800 mm for the wheel diameter, D_{wheel}; 21 for the value of the number of rolling elements, n; 20.5 mm for the rolling elements diameter, d; 167 mm for the pitch diameter, D; 9 ° for the contact angle, ϕ; the range of frequencies considered can be found in table 5.12.

Table 5.12: Range of shaft frequencies considered

train velocity [km/h]	40	50	60	70	80	90	100	110	120
shaft frequency [Hz]	265.258	331.573	397.573	464.202	530.517	596.831	663.146	729.460	795.775

Proceeding to the calculation of the bearing characteristic frequencies, we achieve the following values (see table 5.13):

Table 5.13: Values of bearing characteristic frequencies, in [Hz], for various values of train velocity, in [km/h]

Velocity	40	50	60	70	80	90	100	110	120
BPF	51.761	64.701	77.641	90.581	103.521	116.461	129.401	142.341	155.282
BPFO	41.080	51.350	61.620	71.890	82.160	92.430	102.700	112.970	123.240
FTF	1.956	2.445	2.934	3.423	3.912	4.401	4.890	5.380	5.869
BSF	18.727	23.408	28.090	32.772	37.453	42.135	46.817	51.498	56.180

The SKF industry provides an online range of engineering tools at [46] comprising interactive systems relating to a range of topics associated with rotating machinery, utilizing some of the knowledge that SKF has developed and accumulated. Among the various web tools, the *Frequency Calculator* can be used to calculate the different bearing characteristic frequencies of bearing applications.

Although we have already calculated the pretended frequencies, the utilization of this tool may prove useful to validate the previously obtained values.

Unfortunately, the intended bearing is developed specially for railway applications, and isn't included in the database of the frequency calculator tool. The bearing specifications were input manually (pitch diameter, rolling element diameter, number of rolling elements (per row), contact angle, rotational speed, rotating ring) and the following bearing characteristic frequencies were obtained (see table 5.14).

Table 5.14: Values of bearing characteristic frequencies, in [Hz], for various values of train velocity, in [km/h], obtained in the SKF's frequency calculator tool

Velocity	40	50	60	70	80	90	100	110	120
BPF	51.761	64.701	77.641	90.581	103.521	116.461	129.401	142.341	155.282
BPFO	41.080	51.350	61.620	71.890	82.160	92.430	102.700	112.970	123.240
FTF	1.956	2.445	2.934	3.423	3.912	4.401	4.890	5.380	5.869
BSF	18.727	23.408	28.090	32.772	37.453	42.135	46.817	51.498	56.180

The values obtained are exactly the same that were obtained through the formulation present before, testifying the obtained values. However, it should be remembered that this theoretical approach doesn't account for slip, which always exist in real applications. It should be expected a minimum deviation of 1% between the real values and the ones calculated above.

5.6 Test bench results

5.6.1 Measurement complications

Although the working condition of the test bench had already been checked previously without apparent problems, when the first measurement values were being acquired, the bench started malfunctioning. The shaft was rotating at a high velocities and, although the load applied from the hydraulic jack was low, the rotating shaft suddenly locked in place. The electric motor was stopped, and the test bench condition was checked.

The behaviour of the test bench changed considerably after this event. If no load was applied by the hydraulic jack, the shaft showed a great resistance to rotatory motion, in both direction. However, at the minimum load applied, the shaft could be turned effortlessly. When rotating, at each revolution, the shaft turned faster during the first half of revolution and slowed down during the other half. As the deadline to the submission of this document was near, it wasn't possible to dismount the test bench and find the problem nor replace any faulty components.

The measured values can't be used for outlining the axle-bearing's behaviour nor to extract its features. However, some measurements were still made, and the result will be displayed in the following section. Due to the occurred incident, and to avoid further problems, the values of speed and load chosen were relatively low.

5.6.2 Obtained measurements

To measure the vibration of the test bench, the acquisition equipments was mounted according to the connections specified in section 5.2.1 and 5.2.2. The top of the axle-bearing's box (as the y-axis) and its lateral slot (as the x-axis) were used to fix the accelerometer and measure the vibration values.

Two values for the shaft's rotating speed were considered (displayed directly by the motor's module interface), 265 rpm (corresponding to 40 $\mathrm{km\,h^{-1}}$) and 397 rpm (corresponding to 40 $\mathrm{km\,h^{-1}}$). For each speed, at the lightest load possible, three measurement in each direction were

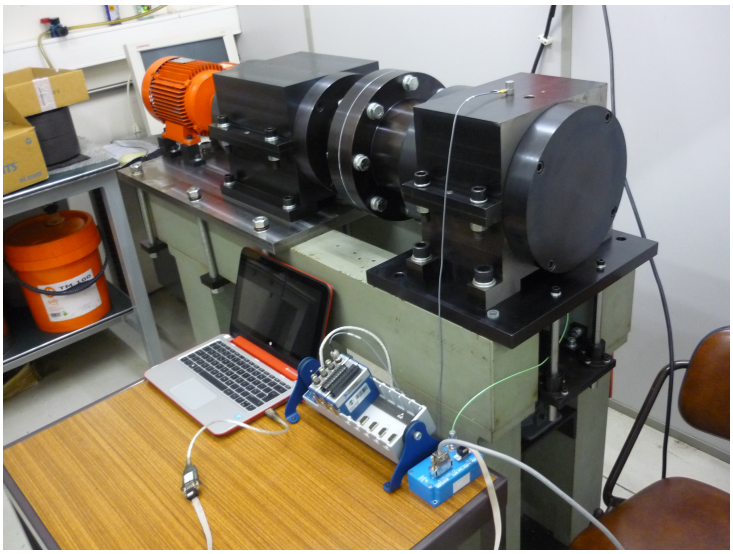


Figure 5.25: Experimental test setup.

made. In the vertical direction, three additional measurements were made, with an applied load of 1 kN. The result were collected through the use of the *signalFFTnoncontinuous.m* and will be displayed in figures 5.26 and 5.27.

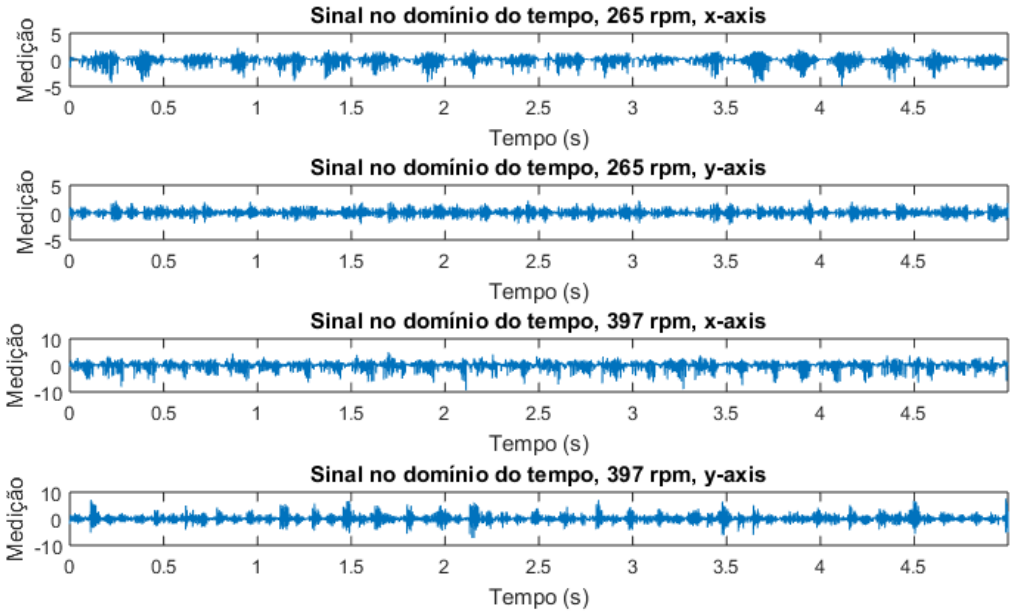


Figure 5.26: Measured values (time-domain) at different speeds.

Through the analysis of figures 5.26 and 5.27, it is clear that a impulsive cyclic feature is present in the signal. It is related to the problem described in section 5.6.1, as as cyclic "knocking" noise coming from the test bench was heard while the results were being collected. As it was already mentioned, and as figure 5.27 shows, when load is applied, the cyclic impulse is less notorious.

Analysing figures 5.28 and 5.29, it is clear that for higher velocities, the vibration signal is more energetic. It is also noticeable that the vibration in the y-direction has higher values than the vibration in the x-direction. This is coherent to the fact that, in the y-direction, the

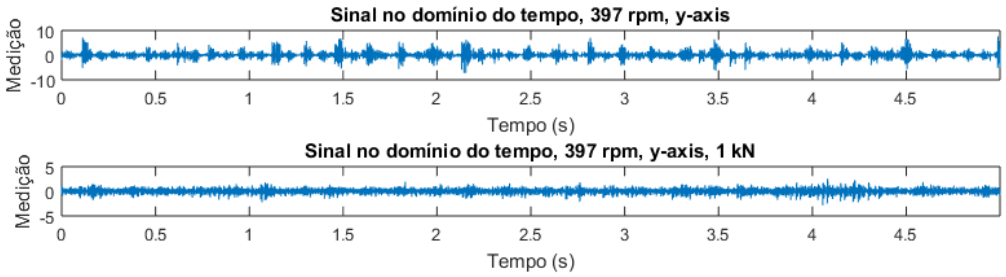


Figure 5.27: Measured values (time-domain), at 397 rpm, for two different loads.

axle-bearing box of the test bench can move freely whereas in the x-direction the movement is limited.

The fact stated above, saying that higher loads lead to a smoother functioning, is once more validated, by the results of figure 5.29.

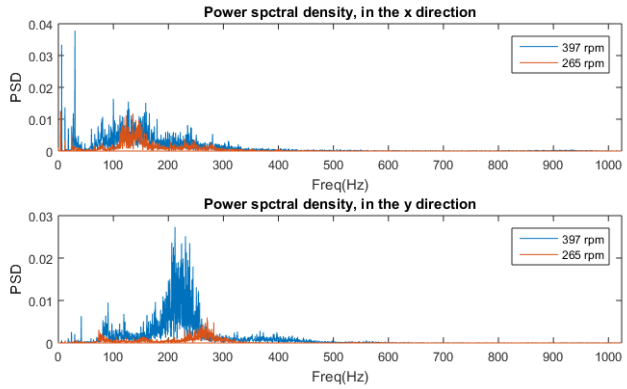


Figure 5.28: Measured values (frequency) at different speeds.

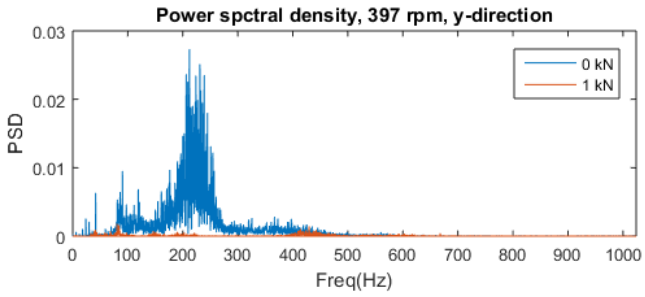


Figure 5.29: Measured values (frequency-domain), at 397 rpm, for two different loads.

Unfortunately, due to the malfunctioning of the test bench, further tests could not be made and an extensive collection and analysis of various results wasn't possible. The evolution of the bearing's vibrational behaviour with different loads and speeds could not be studied. If the test bench was in proper conditions, the following test should be made. The vibrational response of the bearing should be recorded for shaft's speeds equivalent to the vehicle's operating range (40 to 120 $\mathrm{km\,h^{-1}}$). For various speeds, the application of progressive higher loads should also be recorded. The minimum applied load should match the vehicle's load without passenger (47.6 kN), and the maximum one should match its overload (66.5 kN).

5.7 Future alterations to the existing equipment

During the development of the work, several problems were encountered, and some suggestion will be made for their solution:

- The location of the test bench hampered an easy operation, observation and perception of its components. Given the overall size of the test bench, moving it to a wider space would allow a better access and ease the visual analysis of possible problems.
- The validation of the data acquisition system and analysis methods could be firstly done on a simpler and smaller test bench (for smaller bearings), and then proceed to the utilization of the existing one. The confidence on the developed equipment and procedures would assure a faster analysis of the existing bench, since that eventual problems non-related to the test bench would have been previously tracked.
- The existing force sensor was already mounted into the machine, preventing it from being externally tested and its measurements validated. It could be positive in the future to dismount and test it.
- The existing motor calculated the angular velocity through the electric power it consumed. It could be fruitful to mount a tachometer for more trustworthy results.
- For the correction of the problem described in 5.6.1, the test bench should be dismounted and the possible causes searched. It should be also considered if the problem isn't related to the existing bearings (test bearing or supporting bearings). The problem (resistance to movement) worsens as higher speed are achieved by the test bench, and is less notorious when it rests for some time.

Chapter 6

Conclusion

6.1 Conclusions

The use of the right management techniques has a huge impact on any company results. Nowadays, harsher demands are being made at railway companies, for a better, faster and safer service at lower costs. It is crucial to assure an efficient asset management for competitive results.

Railway vehicles have several components that inevitably fail, and rolling element bearings are the ones that have the higher failure rate. However, they are crucial to the vehicles functioning. Their breakdown could lead to catastrophic failure, causalities and high expenses. So, it is important to assure that this component never fails and, at the same time, chose an cost effective maintenance plan.

There are several existing monitoring techniques available, but condition monitoring is the one that assures a best results in almost every existing case. One of the most effective ways to monitor a railway vehicle's axle bearing is through vibration monitoring. Although bearings condition monitoring has received great attention later, it is quite a vast area, and most of techniques and studies were developed only for non-slow speed bearings. So, it is necessary to develop solutions for the study of the applicability of condition monitoring to slow speed bearings. The collection of event data about these bearing's behaviour and its main features is one of the most important steps, since it will assure a solid base for the construction of the remaining work.

During the development of the present work, there were successive problems encountered that hindered a fast evolution of the intended work. Much of the existing material (vibration and force acquisition equipment) had scarce existing information about their utilization, and there was almost no documentation about its previous utilization. There were several constrain concerning the use of the most indicated software (*LabView*, a software created by *National Instruments*, developed to be used with their physic equipment) that forced the use of alternative software (*MATLAB*), with a less direct approach and a longer learning process. The state of the test bench was always unclear. It was build some year ago and hasn't been used since. The test bench's supporting structure had a different behaviour than expected. The existing bearings shouldn't remain unused for a long time (the lubricant present in the bearing degrades quickly if it is not stored properly) and it was probably one of the causes for the test bench malfunctioning. All of the stated facts difficulted the hasty and fruitful progression of the work.

However, some positive results were drawn. Future development is now assured by the certainty of a working acquisition system. It's functioning and properties are detailed in this document, enabling future modification to be done with ease. Some scripts were developed for the calculation of the signal's statistic parameters, the representation of the signal in the time and frequency domain, and for the creation of the signal's kurtogram. However, there is much work to be done, as detailed in the next section.

6.2 Future developments

For future developments, there are several studies that can be done, among others:

- Study and analyse the behaviour and vibrational signal's features of a healthy bearing, for different loads and speeds;
- Simulate different bearing faults on the bearing and analyse it's behaviour and vibrational signal's features, for different loads and speeds;
- Improve the *MATLAB®* code, adding new features to the existing scripts;
- Implement new techniques for data analyses and its automatic application to the acquired data.

Bibliography

[1] Amiya Ranjan Mohanty. *Machinery Condition Monitoring: Principles and Practices*. CRC Press, 2014.

[2] Tedric A Harris and Michael N Kotzalas. *Essential concepts of bearing technology*. CRC press, 2006.

[3] Rolling bearings - skf catalogue, August 2013.

[4] N Symonds, Ilaria Corni, RJK Wood, Adam Wasenczuk, and David Vincent. Observing early stage rail axle bearing damage. *Engineering Failure Analysis*, 56:216–232, 2015.

[5] Idriss El-Thalji and Erkki Jantunen. Dynamic modelling of wear evolution in rolling bearings. *Tribology International*, 84:90–99, 2015.

[6] Gottfried Kuře. Railway technical handbook. axleboxes, wheelset bearings, sensors, condition monitoring, subsystems and services. volume 1, skf. Technical report, ISBN 978-91-978966-3-4, 2011.

[7] Yanping Du, Wenjiao Zhang, Yuan Zhang, Zhenqing Gao, and Xiaohui Wang. Fault diagnosis of rotating machines for rail vehicles based on local mean decomposition—energy moment—directed acyclic graph support vector machine. *Advances in Mechanical Engineering*, 8(1):1687814016629345, 2016.

[8] Wade A Smith and Robert B Randall. Rolling element bearing diagnostics using the case western reserve university data: A benchmark study. *Mechanical Systems and Signal Processing*, 64:100–131, 2015.

[9] Robert B Randall and Jerome Antoni. Rolling element bearing diagnostics—a tutorial. *Mechanical systems and signal processing*, 25(2):485–520, 2011.

[10] Hai Qiu, Jay Lee, Jing Lin, and Gang Yu. Robust performance degradation assessment methods for enhanced rolling element bearing prognostics. *Advanced Engineering Informatics*, 17(3):127–140, 2003.

[11] Simon Iwnicki. *Handbook of railway vehicle dynamics*. CRC press, 2006.

[12] Isao Okamoto. How bogies work. *Railway Technology Today*, 5, 1998.

[13] UPORTO. MAXBE-Interoperable monitoring, diagnosis and maintenace strategies for axle earings. D2.1 - Axle bearing failure modes and degradation process.

[14] Andy Kirwan and Teodor Gradinariu. *Guidelines for the Application of Asset Management in Railway Infrastructure Organisations*. International Union of Railways (UIC), September 2010.

[15] Network Rail. *Asset Management Strategy*, October 2014.

[16] Network Rail. *Asset Management Policy*, March 2014.

[17] Arash Amini, Mani Entezami, and Mayorkinos Papaelias. Onboard detection of railway axle bearing defects using envelope analysis of high frequency acoustic emission signals. *Case Studies in Nondestructive Testing and Evaluation*, 2016.

[18] Keith Bladon, David Rennison, Grigory Izbinsky, Roger Tracy, Trevor Bladon, et al. Predictive condition monitoring of railway rolling stock. *CORE 2004: New Horizons for Rail*, page 22, 2004.

[19] Cai Yi, Jianhui Lin, Weihua Zhang, and Jianming Ding. Faults diagnostics of railway axle bearings based on imf's confidence index algorithm for ensemble emd. *Sensors*, 15(5):10991–11011, 2015.

[20] RW Ngigi, Crinela Pislaru, Andrew Ball, and Fengshou Gu. Modern techniques for condition monitoring of railway vehicle dynamics. In *Journal of Physics: Conference Series*, volume 364, page 012016. IOP Publishing, 2012.

[21] Gottfried Kuře and Víctor Martínez. Applied condition monitoring in railways. http://evolution.skf.com/applied-condition-monitoring-in-railways1/, September 2011.

[22] Jong-Ho Shin and Hong-Bae Jun. On condition based maintenance policy. *Journal of Computational Design and Engineering*, 2(2):119–127, 2015.

[23] Bram de Jonge, Ruud Teunter, and Tiedo Tinga. The influence of practical factors on the benefits of condition-based maintenance over time-based maintenance. *Reliability engineering & system safety*, 158:21–30, 2017.

[24] Khac Tuan Huynh, Anne Barros, Christophe Berenguer, and Inmaculada Torres Castro. A periodic inspection and replacement policy for systems subject to competing failure modes due to degradation and traumatic events. *Reliability Engineering & System Safety*, 96(4):497–508, 2011.

[25] Keomany Bouvard, Samuel Artus, Christophe Berenguer, and Vincent Cocquempot. Condition-based dynamic maintenance operations planning & grouping. application to commercial heavy vehicles. *Reliability Engineering & System Safety*, 96(6):601–610, 2011.

[26] Ranganath Kothamasu, Samuel H Huang, and William H VerDuin. System health monitoring and prognostics–a review of current paradigms and practices. In *Handbook of Maintenance Management and Engineering*, pages 337–362. Springer, 2009.

[27] SW Butcher. Assessment of condition-based maintenance in the department of defense. Technical report, 2000.

[28] J Lee. Approaching zero downtime. *The Center for Intelligent Maintenance Systems, Harbor Research Pervasive Internet Report*, 2003.

[29] Hashem M Hashemian and Wendell C Bean. State-of-the-art predictive maintenance techniques. *IEEE Transactions on Instrumentation and measurement*, 60(10):3480–3492, 2011.

[30] NS Jammu and PK Kankar. A review on prognosis of rolling element bearings. *International Journal of Engineering Science and Technology*, 3(10):7497–7503, 2011.

[31] Sylvester A Aye, P Stephan Heyns, and Coenie JH Thiart. A review of slow speed bearing diagnostics and prognostics. *International Journal of Engineering Science and Technology*, 6(10):726, 2014.

[32] Andrew KS Jardine, Daming Lin, and Dragan Banjevic. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical systems and signal processing*, 20(7):1483–1510, 2006.

[33] Akhand Rai and SH Upadhyay. A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings. *Tribology International*, 96:289–306, 2016.

[34] Abdenour Soualhi, Kamal Medjaher, and Noureddine Zerhouni. Bearing health monitoring based on hilbert–huang transform, support vector machine, and regression. *IEEE Transactions on Instrumentation and Measurement*, 64(1):52–62, 2015.

[35] Andrew C McCormick and Asoke K Nandi. Condition monitoring of rotating machinery using cyclic autoregressive models.

[36] Nader Sawalhi and Robert B Randall. The application of spectral kurtosis to bearing diagnostics.

[37] Diego Alejandro Tobon-Mejia, Kamal Medjaher, Noureddine Zerhouni, and Gerard Tripot. A data-driven failure prognostics method based on mixture of gaussians hidden markov models. *IEEE Transactions on Reliability*, 61(2):491–503, 2012.

[38] Yaguo Lei, Jing Lin, Zhengjia He, and Ming J Zuo. A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mechanical Systems and Signal Processing*, 35(1):108–126, 2013.

[39] Xiao-Sheng Si, Wenbin Wang, Chang-Hua Hu, and Dong-Hua Zhou. Remaining useful life estimation–a review on the statistical data driven approaches. *European journal of operational research*, 213(1):1–14, 2011.

[40] Salem K Al-Arbi. *Condition Monitoring of Gear Systems using Vibration Analysis*. PhD thesis, University of Huddersfield, 2012.

[41] Siyan Wu, Ming J Zuo, and Anand Parey. Simulation of spur gear dynamics and estimation of fault growth. *Journal of Sound and Vibration*, 317(3):608–624, 2008.

[42] Robert Bond Randall. *Vibration-based condition monitoring: industrial, aerospace and automotive applications*. John Wiley & Sons, 2011.

[43] Understanding ffts and windowing. Technical report, National Instruments, december 2016.

[44] Eric Bechhoefer. A quick introduction to bearing envelope analysis, 2016.

[45] Jerome Antoni. https://www.mathworks.com/matlabcentral/fileexchange/48912-fast-kurtogram, December 2014.

[46] Engineering tools. http://www.skf.com/group/knowledge-centre/engineering-tools/index.html.

Appendix A

Axle bearing

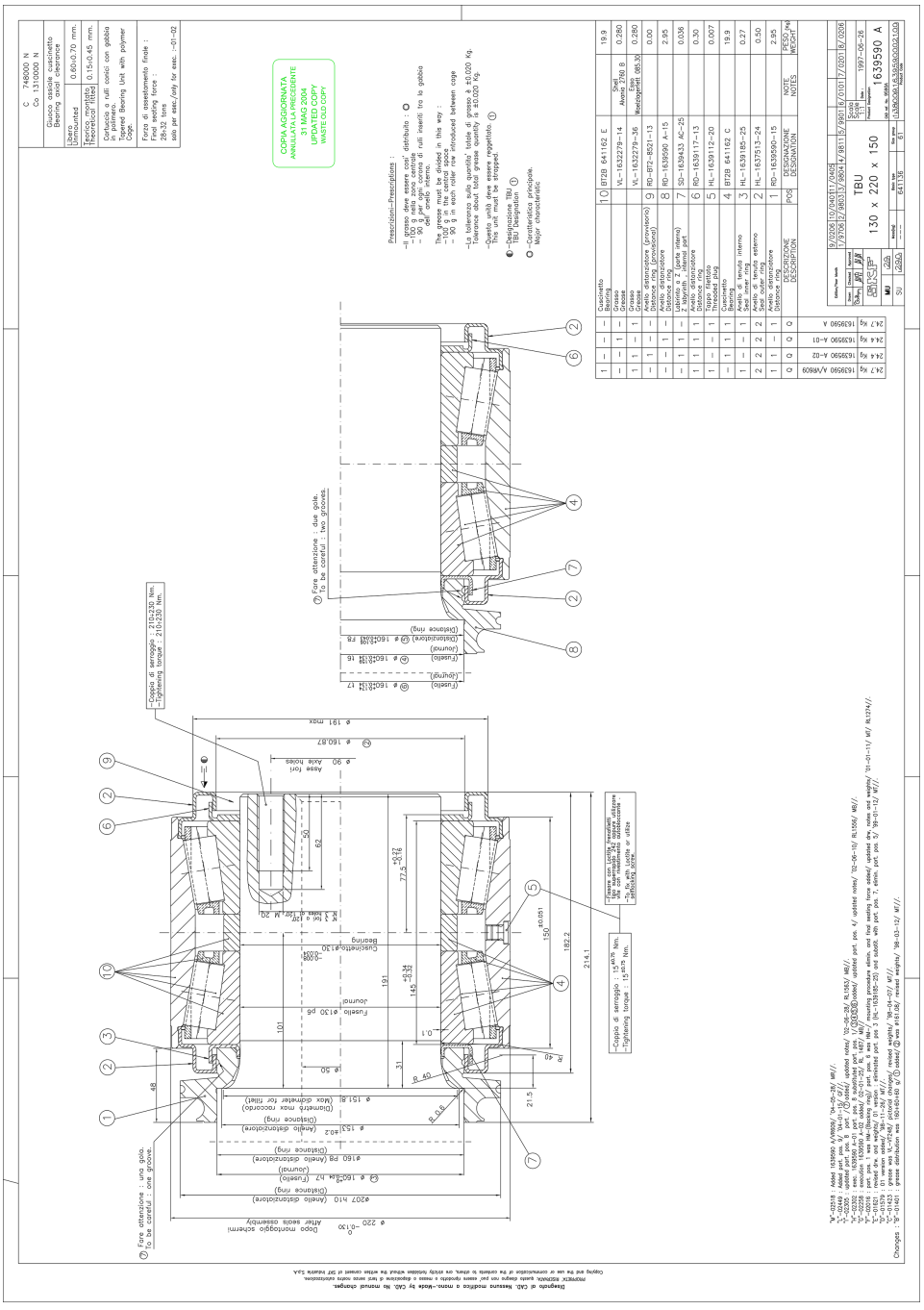


Figure A.1: Technical drawing of the *TBU 1639590 A* bearing.

Appendix B

UTE 2240

Série 2240 - Ficha Técnica

Número de Série: 2241-2297
Número de Unidades: 57
Ano de Entrada ao Serviço: 2003

Tipo de Tracção: Eléctrica
Natureza do Serviço: Regional e Intercidades
Bitola de Via: 1668 mm

Características Técnicas

Construtores:

Caixa	Sorefame
Equip. Eléctrico	Alstom
Transmissão	Alstom
Freio Pneumático	Knorr

Características de Funcionamento:

Tracção:

Velocidade Máxima	120 Km/h
Esforço de Tracção no Arranque	106 kN
Esforço de Tracção à Velocidade Máx.	23 kN

Freio Dinâmico:

Esforço Max. nas Rodas	75 kN
Vel. Correspondente	entre 10 e 90 kms/h
Esforço à Vel Máxima	50 kN

Características Gerais:

Tipo da Automotora	U.T.E.
Potência Nominal (rodas)	960 kW
Disposição dos Rodados	2' 2' + Bo' Bo' + 2' 2'
Diâmetro das Rodas (novas)	1000 mm (Motora) 850 mm (Reboques)
Número de Cabinas de Condução	2
Freio Dinâmico	Eléctrico por recuperação
Areeiros (número)	Não Tem
Sistema de Homem-Morto	Alstom
Comando em Unidades Múltiplas	até 3
Lubrificador de Verdugos	Não Tem
Registador de Velocidade	CONVEL + Deuta-Werke
Rádio Solo-Comboio	ASCOM / NEC
Controlo Automático de Velocidade	Bombardier

Dimensões e Capacidade

Pesos e Capacidade:

Lotação (Lugares em Pé):

Carga Normal (3 passageiros / m2)	163
Carga Máxima (5 passageiros / m2)	272
Sobrecarga (corredores: 5 pass / m2 + vestíbulos: 7 pass / m2)	308

Lotação (Lugares Sentados):

Motora	106
Reboque	79
Total (U.T.E.)	264

Massas:

Motor de Tracção	800 Kg
Transformador Principal	3.410 Kg
Conversor Auxiliar	1.750 Kg
Conversor de Tracção	2.340 Kg
Sistema de Ar Condicionado	1.034 Kg
Bogie (Motor)	10.470 Kg
Bogie (Livre)	5.865 Kg

Carga (70 Kg/Passageiro):

Normal	29,9 t
Máxima	37,5 t
Sobrecarga	40,0 t
Tara em Ordem de Marcha	135,7 t

Equipamento Eléctrico

Motores de Tracção:

Construtor	Alstom
Modelo	4 EXA 2146B
Características	Trifásico Assincrono 4 Polos Autoventilado
Potência em Regime Contínuo	355 kW @ 2.200 rpm
Tensão (por fase e composta)	760 / 1.360 V
Corrente	166 A
Frequência	74,5 Hz
Gama de Velocidades	0 a 3.648 rpm

Pantógrafo:

Fabricante	Schunk
Modelo	WBL85
Tipo de Accionamento	Pneumático (5.5 bar)
Paleta	Escovas em Paralelo Suspensão Independente
Velocidade Máxima	160 km/h
Material das Escovas	Grafite
Tensão Nominal	25 kV AC
Corrente Nominal	350 A
Corrente Máxima	600 A

Figure B.1: Technical drawing of UTE 2240 (part 1)

Conversor de Tracção:

Fabricante	Alstom
Referencia	10SF189B2
Tensão Nominal de Alimentação	900 V a.c.
Gama de Tensões de saída do Ondulador	0 a 1.160 V
Gama de Frequências de saída do Ondulador	0 a 150 Hz
Frequência de Comutação dos IGTB's do Ondulador	600 Hz
Potência Nominal do Ondulador	500 kW

Baterias:

Tipo	Ácidas
Fabricante	HOPPECKE
Número de Elementos	36
Tensão Nominal	72 V c.c.
Capacidade	180 A/h

Transformador Principal:

Fabricante	Pauwels
Modelo	T-25KVM
Tensão do Primário	25 kV
Corrente do Primário	50 A
Potência Nominal do Primário	1.250 kVA
Frequência	50 Hz +- 1 Hz
Número de Fases	1
Tensões dos Secundários	2 x 900 V
Correntes dos Secundários	2 x 694 A
Potência Nominal dos Secundários	2 x 625 kVA

Conversor Auxiliar:

Fabricante	Alstom
Gama da Tensão de Alimentação	1.200 a 2.200 V d.c.
Tensão Nominal de Saída (AC)	400 V +- 5%
Frequência (AC)	50 Hz +- 2%
Potência Nominal (AC)	140 kVA
Potência de Pico (AC)	250 kVA
Tensão Nominal de Saida (DC)	72 V d.c.
Potência Nominal (DC)	15 kW

Conforto e Informação aos Passageiros

Características dos Salões:

Caudal de ar exterior	1470 m3/h
Caudal de ar de retorno	2815 m3/h
Caudal de ar tratado	4285 m3/h
Potência de refrigeração	46,2 kW
Potência de aquecimento	30 kW

Características das Cabines:

Caudal de ar exterior	60 m3/h
Caudal de ar de retorno	700 m3/h
Caudal de ar tratado	640 m3/h
Potência de refrigeração	4 kW
Potência de aquecimento	4 kW

Equipamento de Climatização:

Fabricante	MERAK
Fluido refrigerante	R-407C

Instalação Sonora:

Fabricante	Alstom
Funcionalidades	Mensagens pré-gravadas Música ambiente com controlo de volume Regime de Paragens

Videovigilância:

Fabricante	Petards - Joyce Loebl
Câmaras	3 a cores em cada veículo
Gravador	Digital com registo de aprox. 60 horas

Outro Equipamento

Produção de Ar Comprimido:

Compressor	Rotativo de Parafuso
Fabricante	Knorr
Tipo	SL22-5-77
Pressão Nominal	10 bar
Velocidade de Rotação	1.470 r.p.m.
Caudal	950 l/min

Portas:

Accionamento	Eléctrico c/ Anti-Entalamento
Número (por lateral)	5
Largura Útil	1.020 mm
Altura do Piso	1.290 mm

Engates:

Extremos	Automático DELLNER (tipo Scharfenberg)
Intermédios	Semi-Permanente (Scharfenberg Schaku)

Figure B.2: Technical drawing of UTE 2240 (parte 2)

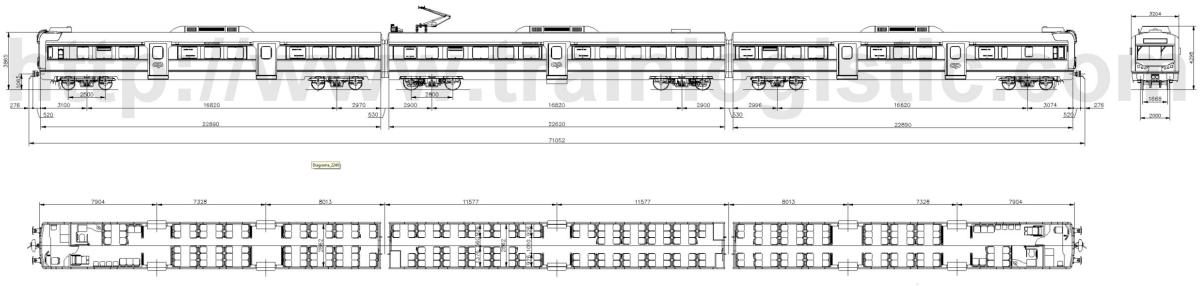


Figure B.3: Technical drawing of UTE 2240

Appendix C

MATLAB code

signalFFTcontinuous.m

```
display('Started')
clearvars
clear functionplotData.m

%==========DEFINIÇÕES MEDIÇÃO============
slot=1; %posição da placa no chassi
channel=0; %canal da placa de aquisição
mtype='Accelerometer'; %tipo de medição Voltage/Accelerometer/IEPE

rate=2048; %taxa de amostragem
sensitivity=0.01;%[V/G]
excicurr=0.002;%[V] corrente de excitação, por predefinição 0.002V

width=2; %largura máxima do gráfico da medição, em segundos
%ESCOLHER APENAS 1 das seguintes (colocar % na outra):
secref=2; %actualiza gráfico a cada X segundos
%ou
%scaref=2048; %actualiza gráfico a cada X medições
%se nenhuma for seleccionada, o valor será ajustado automáticamente para
%actualizar 10 vezes por segundo

%==================END=================

if (exist('secref','var')==1 && exist('scaref','var')==1)
    display(sprintf('\nApenas a uma das seguintes variáveis deve ser usada:\n ...
         "secref" ou "scaref"\ncolocar "%%" atrás da outra, ou de ambas para ...
        valor automático\n'))
    return
end

display(sprintf ...
    ('=====================================================================\nPara ...
    eventuais problema, executar o programa\n   "NI MAX - Measurement & ...
    Automatic Explorer"\ne verificar se as placas são devidamente ...
    indentificadas.\nSe mesmo assim se mantiver, executar a acção:\n   ...
    "daqreset"\n(ou "DataAcquisitionToolApp" e clicar File>Rescan for ...
    Devices)\n======================================================================'))

d=daq.getDevices;

global s;
s=daq.createSession('ni');

```

```
s.IsContinuous=true;

%Preparação variáveis
global lh;
global lim;
global fig;
global datasignal;
global datat;
global datafreq;
global datafft;

s.Rate=rate;
pot=nextpow2(lim);
for i=1:size((d),2)
if d(i).SlotNumber==slot
slotID=i;
end
end
channel=strcat('ai',num2str(channel));
if exist('secref','var')==1
    s.IsNotifyWhenDataAvailableExceedsAuto=false;
    s.NotifyWhenDataAvailableExceeds=round(rate*secref);
else
    if exist('scaref','var')==1
        s.IsNotifyWhenDataAvailableExceedsAuto=false;
        s.NotifyWhenDataAvailableExceeds=scaref;
    end
end
%end

ch=addAnalogInputChannel(s,d(slotID).ID,channel,mtype);
if strcmp(mtype,'IEPE')==1
ch.ExcitationCurrent=excicurr;
end
if strcmp(mtype,'Accelerometer')==1
ch.Sensitivity=sensitivity;
ch.ExcitationCurrent=excicurr;
end

lh = s.addlistener('DataAvailable',@functionplotData);

fig=figure('Name','Sinal medido','NumberTitle','off');
subplot(2,1,1);
title('Sinal no domínio do tempo')
ylabel('Medição')
xlabel('Tempo (s)')

subplot(2,1,2);
title('Sinal no domínio da frequência (FFT) - Power Spectral Density')
ylabel('Medição')
xlabel('Frequência (Hz)')

pushbutton=uicontrol('Style', 'pushbutton', 'String', 'Pause', 'Callback', ...
    @functionbutton);


display(sprintf('\nStarted acquisition\n'))

s.startBackground();
lim=round(s.Rate*width);

%resto do código na função functionplotData.m

```

```
display(sprintf('\nEnd\n--------'))
```

signalFFTnoncontinuous.m

```
display('Started')
clearvars


%DEFINIÇÕES MEDIÇÃO
slot=1; %posição da placa no chassi
channel=0; %canal da placa de aquisição
mtype='Accelerometer'; %tipo de medição Voltage/Accelerometer/IEPE

rate=2048; %taxa de amostragem
sensitivity=10.14E-3;%[V/G]
excicurr=0.002;%[V] corrente de excitação, por predefinição 0.002V


%ESCOLHER APENAS 1 das seguintes (colocar % na outra):
%nscans=1050; %número de amostras
%ou
duration=3; %duração da medição
%=====================================================


if exist('nscans','var')==1 && exist('duration','var')==1
    display(sprintf('\nApenas a uma das seguintes variáveis deve ser ...
        atribuido um valor:\n  "nscans" ou "duration"\ncolocar "%%" atrás da ...
        outra\n'))
    return
end

display(sprintf ...
    ('=====================================================================\nPara ...
    eventuais problema, executar o programa\n   "NI MAX - Measurement & ...
    Automatic Explorer"\ne verificar se as placas são devidamente ...
    indentificadas.\nSe mesmo assim se mantiver, executar a acção:\n   ...
    "daqreset"\n(ou "DataAcquisitionToolApp" e clicar File>Rescan for ...
    Devices)\n======================================================================'))

d=daq.getDevices;

s=daq.createSession('ni');
%Preparação variávei
for i=1:size((d),2)
if d(i).SlotNumber==slot
slotID=i;
end
end
channel=strcat('ai',num2str(channel));

s.Rate=rate;
if exist('nscans','var')==1
    s.NumberOfScans=nscans;
else
    s.DurationInSeconds=duration;
end
%end

ch=addAnalogInputChannel(s,d(slotID).ID,channel,mtype);
if strcmp(mtype,'IEPE')==1
```

```
ch.ExcitationCurrent=excicurr;
end
if strcmp(mtype,'Accelerometer')==1
ch.Sensitivity=sensitivity;
ch.ExcitationCurrent=excicurr;
end

display(sprintf('\nStarted acquisition\n'))

[result,t]=s.startForeground();

display(sprintf('\nStoped acquisition\n'))

if strcmp(mtype,'IEPE')==1
result=result/sensitivity;
end

pot=nextpow2(size((result),1));
if (2^pot)==size((result),1)
else
    pot=pot-1;
    formatSpec = char('ATENÇÃO:\nPara a passagem para o domínio da frequência ...
        através de FFT é aconselhável que o número de pontos\nseja uma ...
        potência de 2.\nDos %d valores recolhidos, serão apenas usados os ...
        primeiros %d.\nPara que não sejam ignorados valores, dever-se-á ...
        aumentar o valor total de amostras para %d.');
    display(sprintf(formatSpec,size((result),1),(2^pot),(2^(pot+1))))
    trigger=1;
end

resulthanning=result(1:2^pot).*hanning(2^pot);
resultffthanning=2*fft(resulthanning,2^pot);%o 2* é por causa do hanning
resultpsdhanning=resultffthanning.*conj(resultffthanning)/(2^pot);
yvalshanning=resultpsdhanning(1:(2^pot/2+1));

xvals=(s.Rate*(0:(2^pot-1))/(2^pot))';
xvals=xvals(1:2^pot/2+1);


figure('Name','Sinal medido','NumberTitle','off')
subplot(2,1,1);
plot(t,result)
xlim([0 t(end,1)])
title('Sinal no domínio do tempo')
ylabel('Medição')
xlabel('Tempo (s)')

subplot(2,1,2);
plot(xvals,yvalshanning)
xlim([0 xvals(end,1)])
if exist('trigger','var')==1
    title({'Sinal no domínio da frequência (FFT) - Power Spectral ...
        Density',['(considerando apenas os primeiros',' ',num2str(2^pot),' ...
        valores medidos)']})
else
    title('Sinal no domínio da frequência (FFT) - Power Spectral Density')
end
ylabel('Medição')
xlabel('Frequência (Hz)')

%subplot(3,1,3);
%plot(t,result.*hanning(2^pot))

```

```
107 stop(s)
108
109 statN=length(result);
110 statxmedio=mean(result);
111 statdif=(result-statxmedio);
112
113 %RMS:
114 rmsvalue=sqrt(sum(statdif.^2)/statN); %função matlab: rmsvalue2=rms(result)
115
116 %Kurtosis:
117 kurt=(sum(statdif.^4)/statN)/(sum(statdif.^2)/statN)^2; %função matlab: ...
        kurt2=kurtosis(result)
118
119 %crestfactor
120 crestfactor=max(result)/rmsvalue; %função matlab: crestfactor2=peak2rms(result)
121
122 %skewness
123 skew=(sum(statdif.^3)/statN)/rmsvalue^3; %função matlab: skew2=skewness(result)
124
125 formatSpec = char('\nValores dos parametros estatisticos:\n   RMS: %d\n   ...
        Kurtosis: %d\n   Crest factor: %d\n   Skewness: %d\n');
126 display(sprintf(formatSpec,rmsvalue,kurt,crestfactor,skew))
127
128 display(sprintf('\nEnd\n--------'))
```

functionplotData.m

```
1  function plotData(src,event)
2
3      global lim;
4      global fig;
5      global s;
6      global datasignal;
7      global datat;
8      global datafreq;
9      global datafft;
10
11
12     persistent t;
13     persistent result;
14     persistent xvals;
15     persistent yvals;
16
17     if (isempty(t))
18         t = [];
19         result = [];
20         xvals = [];
21         yvals = [];
22     end
23
24     t=[t;event.TimeStamps];
25     result=[result;event.Data];
26
27     if size((result),1)<=lim
28         figure(fig)
29         subplot(2,1,1);
30         plot(t,result);
31         xlim([t(1,1) t(end,1)])
32     else
33         figure(fig)
34         subplot(2,1,1);
```

```
35          plot(t((end-lim):end,1),result((end-lim):end,1));
36          xlim([t(end-lim,1) t(end,1)]);
37      end
38
39      if size((result),1)<lim
40          pot=nextpow2(size(result,1));
41          if (2^pot)~=(size(result,1))
42              pot=pot-1;
43          end
44          aux=result(end-2^pot+1:end,1);
45          resulthanning=aux.*hanning(2^pot);
46          resultfffthanning=2*fft(resulthanning);%o 2* é por causa do hanning
47          resultpsdhanning=resultfffthanning.*conj(resultfffthanning)/(2^pot*s.Rate);
48          yvalsaux=resultpsdhanning(2:(2^pot/2+1));
49          xvalsaux=(s.Rate*(0:(2^pot-1))/(2.^pot))';
50          xvalsaux=xvalsaux(2:2^pot/2+1);
51          figure(fig)
52          subplot(2,1,2);
53          semilogy(xvalsaux,yvalsaux);
54          xlim([xvalsaux(1,1) xvalsaux(end,1)]);
55          %parametros estatísticos
56          statN=length(aux);
57          statxmedio=mean(aux); %=sum(a)/length(a)
58          statdif=(aux-statxmedio);
59          rmsvalue=sqrt(sum(statdif.^2)/statN);
60          kurt=(sum(statdif.^4)/statN)/(sum(statdif.^2)/statN)^2;
61          crestfactor=max(aux)/rmsvalue;
62          skew=(sum(statdif.^3)/statN)/rmsvalue^3;
63          inf=strcat('RMS: ',num2str(rmsvalue),'   Crest factor: ...
                ',num2str(crestfactor),'   Skewness: ',num2str(skew),'   ...
                Kurtosis:', num2str(kurt));
64          title(sprintf('Sinal no domínio da frequência (FFT) - Power Spectral ...
                Density \n %s',info))
65      else
66          pot=nextpow2(lim);
67          if 2^pot~=lim
68              pot=pot-1;
69          end
70          aux=result(end-2^pot+1:end,1);
71          resulthanning=aux.*hanning(2^pot);
72          resultfffthanning=2*fft(resulthanning);%o 2* é por causa do hanning
73          resultpsdhanning=resultfffthanning.*conj(resultfffthanning)/(2^pot*s.Rate);
74          yvalsaux=resultpsdhanning(2:(2^pot/2+1));
75          xvalsaux=(s.Rate*(0:(2^pot-1))/(2.^pot))';
76          xvalsaux=xvalsaux(2:2^pot/2+1);
77          figure(fig)
78          subplot(2,1,2);
79          semilogy(xvalsaux,yvalsaux);
80          xlim([xvalsaux(1,1) xvalsaux(end,1)]);
81          %parametros estatísticos
82          statN=length(aux);
83          statxmedio=mean(aux); %=sum(a)/length(a)
84          statdif=(aux-statxmedio);
85          rmsvalue=sqrt(sum(statdif.^2)/statN);
86          kurt=(sum(statdif.^4)/statN)/(sum(statdif.^2)/statN)^2;
87          crestfactor=max(aux)/rmsvalue;
88          skew=(sum(statdif.^3)/statN)/rmsvalue^3;
89          inf=strcat('RMS: ',num2str(rmsvalue),'   Crest factor: ...
                ',num2str(crestfactor),'   Skewness: ',num2str(skew),'   ...
                Kurtosis:', num2str(kurt));
90          title(sprintf('Sinal no domínio da frequência (FFT) - Power Spectral ...
                Density \n %s',info))
91      end
```

```
92
93      if size(xvalsaux,1)<lim
94      c=lim-size(xvalsaux,1);
95      xvalsaux=[xvalsaux;zeros(c,1)];
96      yvalsaux=[yvalsaux;zeros(c,1)];
97      end
98
99
100
101     xvals=[xvals xvalsaux];
102     yvals=[yvals yvalsaux];
103     datasignal=result;
104     datat=t;
105     datafreq=xvals;
106     datafft=yvals;
107
108 end
```

SpecKurt.m

```
1  v=result;
2  sr=rate;
3
4  %Outputs:
5  % S :Kurtogram
6
7  %Definition of general settings
8  lvl = [2 3 4 6 8 12 16 32]; %Octave Levles
9  lw = 10;%lowerest window frequency value
10 hi = sr/2; %Nyquest
11 n = length(lvl);
12 S = zeros(n,32);
13
14 %Defenition for the calculation of Spectrum
15 winlen= 128;% windln :window length, e.g. 2048,
16 overlen= 64;% noverlap :length of the overlp
17
18 %END%
19
20 for i = 1:n,
21     cLvl = lvl(i);
22     dBw = (hi-lw)/cLvl;
23     cLw = lw;
24     dix = floor(32/cLvl);
25     sdix = 1;
26     edix = dix;
27     for j = 1:cLvl,
28         cHi = cLw + dBw;
29         [y,dt] = envelope(v,1/sr,cLw,cHi);
30         [P,frq] = psde(y,winlen,1/dt,overlen);
31         sk = kurtosis(P);
32
33         idx = sdix:edix;
34         S(i,idx) = sk;
35         sdix = edix + 1;
36         edix = edix + dix;
37         cLw = cLw + dBw;
38     end
39 end
40 fq = linspace(lw,hi,32);
41 surf(fq(1:end-1),1:8,S(:,1:end-1))
```

```
axis([fq(1) fq(end-1) 1 8])
% title(name);
colorbar
```

envelope.m

```
function [env,dty] = envelope(data,dt,lowf,highf)
% [env,dty] = envelope1(data,dt,nfilt,lowf,highf);
%Inputs:
% data :data vector, time domain
% dt :sampling time interval
% lowf :low frequency limit of bandpass filter
% highf :high frequency limit of bandpass filter
%Outputs:
% env :Envelope of data
% dty :decimated sample rate
n = length(data);
dfq = 1/dt/n;
idxLow = floor(lowf/dfq);
idxHi = ceil(highf/dfq);
D = fft(data);
idx = idxHi-idxLow + 1;

D(1:idx) = D(idxLow:idxHi);
D(idx+1:end) = 0;
data = abs(ifft(D));
bw = highf - lowf;
r = fix(1/(bw*2*dt));
env = data(1:r:n); % crop first nfilt elements

dty = dt*r;
```

psde.m

```
function [Spec, freq] = psde(x, winln,Fs, noverlap)
%[Spec, freq] = psde(x, winln,Fs, noverlap);
%Inputs:
% x :time domain data
% windln :window length, e.g. 2048,
% Fs :sampleing frequency
% noverlap :length of the overlp
%Outputs:
% Spec :vector Spectrum
% freq :corresponding frequency

n = winln;
m = n/2;
window = .5*(1 - cos(2*pi*(1:m)'/(n+1))); %Hann window
window = [window; window(end:-1:1)];
window = window(:);

nfft = length(window);

n = length(x); % Number of data points
nwind = length(window); % length of window
if n < nwind % zero-pad x if it has length less than the window length
    x(nwind)=0; n=nwind;
```

```
24 end
25 % Make sure x is a column vector; do this AFTER the zero-padding
26 % in case x is a scalar.
27 x = x(:);
28
29 k = fix((n-noverlap)/(nwind-noverlap)); % Number of windows
30 index = 1:nwind;
31 %KMU = norm(k*window)^2; % Normalizing scale factor ==>asymptotically unbiased
32 KMU = k*sum(window)^2;% alt. Nrmlzng scale factor ==> peaks areabout right
33
34 Spec = zeros(nfft,1);
35 for i=1:k
36     xw = window.*detrend(x(index));
37     index = index + (nwind - noverlap);
38     xx = fft(xw,nfft);
39     Xx = abs(xx).^2;
40     Spec = Spec + Xx;
41 end
42
43 % Select first half
44 if ~any(any(imag(x)~=0)), % if x is not complex
45     if rem(nfft,2), % nfft odd
46         select = (1:(nfft+1)/2)';
47     else
48         select = (1:nfft/2)';
49     end
50     Spec = Spec(select);
51 else
52     select = (1:nfft)';
53 end
54
55 freq = (select - 1)*Fs/nfft;
56
57 % find confidence interval if needed
58
59 Spec = sqrt(Spec*(4/KMU)); % normalize: ow in Energy vs. Power
```